(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 September 2005 (09.09.2005)

PCT

(10) International Publication Number
WO 2005/081716 A2

(51) International Patent Classification: Not classified

(21) International Application Number: PCT/US2004/039579

(22) International Filing Date: 24 November 2004 (24.11.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/524,479 24 November 2003 (24.11.2003) US

(71) Applicant (*for all designated States except US*): **THE JOHNS HOPKINS UNIVERSITY** [US/US]; 100 North Charles Street, Baltimore, MD (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **WU, Tzyy-Choou** [US/US]; 11002 Nacirema Lane, Stevenson, MD 21153 (US). **HUNG, Chien-Fu** [—/US]; 13 Ballyhean Court, Timonium, MD 21093 (US). **KIM, Tae Woo** [KR/KR]; Korea Research Institute of Chemical Technology, P.O.Box 107, Yuseon-gu, Daejeon, 305-600, Korea (KR).

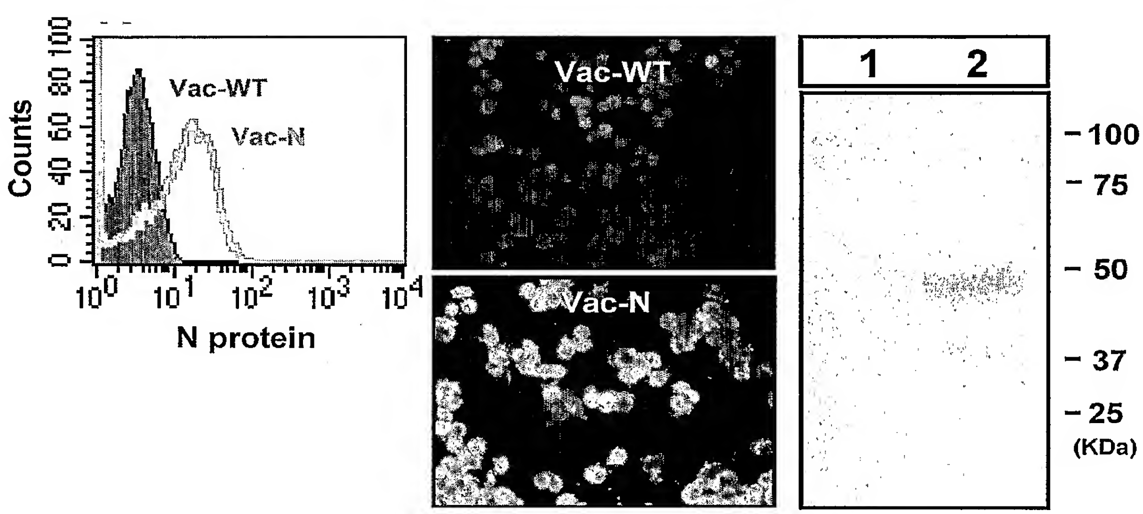
(74) Agent: **SHMUEL, Livnat**; McKenna Long & Aldridge LLP, 1900 K Street , N.W., Washington, DC 20006 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,

[Continued on next page]

(54) Title: DNA VACCINES TARGETING ANTIGENS OF THE SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS (SARS-CoV)



(57) Abstract: This invention provides compositions and methods for inducing and enhancing immune responses, particularly antigen-specific CD8+ T cell mediated responses, against antigens of the SARS coronavirus. These antigens include epitopes of the Membrane (M), Envelope (E), Spike (S) and Nucleocapsid (N) proteins of the virus. Such responses are induced using DNA constructs as an immunogens or vaccines, which encode chimeric polypeptides comprising endoplasmic reticulum chaperone polypeptides, such as human calreticulin (CRT) and an antigenic peptide or polypeptide. In particular, the invention provides compositions and methods for enhancing immune responses induced by polypeptides made *in vivo* by administered nucleic acid, such as naked DNA or expression vectors, encoding the chimeric molecules. Such enhanced immunity, whether T cell mediated or antibody-mediated protects an infected subject from infection or spread of the SARS CoV *in vivo*.

WO 2005/081716 A2

FR, GB, GR, HU, IE, IS, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DNA Vaccines Targeting Antigens of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV)

BACKGROUND OF THE INVENTION

Field of the Invention

This invention, in the field of immunology, virology and medicine, provides immunogenic compositions and methods for inducing enhanced antigen-specific immune responses, particularly those mediated by cytotoxic T lymphocytes (CTL), using chimeric or hybrid nucleic acid molecules that encode an endoplasmic reticulum chaperone polypeptide, *e.g.*, calreticulin, and a polypeptide or peptide antigen of the SARS coronavirus (SARS-CoV).

Description of the Background Art

DNA vaccines are known for their ability to induce both cellular and humoral antigen-specific immunity (reviewed in Donnelly, J *et al.*, 1997. *Annu Rev Immunol 15*:617-648 ; Robinson, HL, 1997. *Vaccine 15*:785-787; Sin, JI *et al.* 2000, *Intervirology 43*:233-246). Advantages of DNA is that it is relatively stable, and it can be easily prepared and harvested in large quantities. In addition, naked plasmid DNA is relatively safe and therefore can be repeatedly administered as a vaccine (Donnelly *et al., supra*; Robinson, *supra*). However, naked DNA lacks cell targeting specificity making it important to find an efficient route for delivery into appropriate target cells, such as professional antigen-presenting cells (APCs). Intradermal (i.d.) administration of DNA immunogens or vaccines using a gene gun represents a convenient form of delivery to professional APCs, such as dendritic cells (DCs), *in vivo* (Condon, C *et al.*, 1996, *Nat Med 2*:1122-8). DCs are the most potent professional APCs for priming CD4+ T helper and CD8+ T cytotoxic or killer T cells *in vivo* (reviewed in Cella, M *et al.*, 1997, *Curr Opin Immunol 9*:10-16; Hart, DN, 1997, *Blood 90*:3245-3287; Steinman, RM, 1991, *Annu Rev Immunol 9:*271-296). Thus, gene gun delivery of DNA vaccines to DCs has become an important method for enhancing T cell-mediated immunity against viral infection.

Forms of DNA vaccines include "naked" DNA, such as plasmid DNA (U.S. Patent Nos. 5,580,859; 5,589,466; 5,703,055), viral DNA, and the like. Basically, a DNA molecule encoding a desired immunogenic protein or peptide is administered to an individual and the protein is generated *in vivo*. Use of "naked" DNA vaccines has the advantages of being safe because, *e.g.*, the plasmid itself has low immunogenicity, it can be easily prepared with high purity and, compared to proteins or other biological reagents, it is highly stable. However, DNA vaccines have limited potency. Several strategies have been applied to increase the potency of

DNA vaccines, including, *e.g.*, targeting antigens for rapid intracellular degradation; directing antigens to APCs by fusion to ligands for APC receptors; fusing antigens to chemokines or to antigenic pathogenic sequences, co-injection with cytokines or co-stimulatory molecules or adjuvant compositions.

Antiviral and antitumor vaccines are an attractive approach for treatment of viral illnesses and cancer because they may have the potency to eradicate systemic virus (or virus-infected cells) or tumor cells in multiple sites in the body and the specificity to discriminate between neoplastic and non-neoplastic cells (Pardoll (1998) *Nature Med. 4*:525-531). Effective anti-viral and most anti-tumor effects of the immune system are mediated by cellular immunity. The cell-mediated component of the immune system is equipped with multiple effector mechanisms capable of eradicating virus-infected cells and tumors, and most of these responses are regulated by T cells. Therefore, there is a need in the art for antiviral or anticancer vaccines, particularly as DNA vaccines, that enhance virus-specific (or tumor-specific) T cell responses, to treat virus infections and to control tumors.

HPV oncogenic proteins, E6 and E7, are co-expressed in most cervical cancers associated with HPV and are important in the induction and maintenance of cellular transformation. Therefore, in earlier studies, the present inventors and colleagues have described nucleic acid vaccines targeting E6 or E7 proteins as an approach to prevent and treat HPV-associated cervical malignancies. HPV-16 E7 and E6 are a well-characterized cytoplasmic/nuclear proteins.

Calreticulin and Related Proteins

Calreticulin (CRT), an abundant 46 kilodalton (kDa) protein located in the lumen of the cell's endoplasmic reticulum (ER), displays lectin activity and participates in the folding and assembly of nascent glycoproteins. See, *e.g.*,, Nash (1994) *Mol. Cell. Biochem. 135*:71-78; Hebert (1997) *J. Cell Biol. 139*:613-623; Vassilakos (1998) *Biochemistry 37*:3480-3490; Spiro (1996) *J. Biol. Chem. 271*:11588-11594; Conway, EM *et al.*, 1995. Heat shock-sensitive expression of calreticulin. *In vitro* and *in vivo* up-regulation. *J Biol Chem 270*:17011-17016) CRT is related to the family of heat shock proteins (HSPs) (Basu, S. *et al.*, *J. Exp. Med. 189*:797-802; Conway *et al.*, *supra)* and associates with peptides transported into the ER by transporters that are involved in antigen processing, such as TAP-1 and TAP-2 (Spee *et al.*, (1997) *Eur. J. Immunol. 27*:2441-2449) and with MHC class I-β2m molecules to aid in antigen presentation Sadasivan, B *et al.*, 1996, *Immunity 5*:103-114; CRT also forms complexes with

peptides *in vitro*. Upon administration to mice, such peptide-CRT complexes, elicited peptide-specific CD8+ T cell responses (Basu *et al., supra*; Nair, 1999, *J. Immunol. 162*:6426-6432). CRT purified from murine tumors elicited immunity specific for the tumor from which the CRT was taken, but not for an antigenically distinct tumor (Basu, *supra*). By pulsing mouse dendritic cells (DCs) *in vitro* with a CRT-peptide complex, the peptide was re-presented by MHC class I molecules on the DCs to stimulate a peptide-specific CTL response(Nair, *supra*).

The present inventors and their colleagues have previously used the approach of fusing or combining, at the DNA (or RNA) level, a nucleotide sequence encoding an antigen to test several intracellular targeting strategies that enhance MHC class I and/or class II processing and antigen presentation (Hung, CF. *et al.*, 2003, Improving DNA vaccine potency via modification of professional antigen presenting cells. *Curr Opin Mol Ther 5:20-24*. Recently, several of the present inventors performed direct comparisons of these strategies for their ability to improve DNA vaccine potency. This comparison showed that linkage of antigen to CRT in a DNA vaccine resulted in the most marked enhancement of the humoral and T cell-mediated immune responses in vaccinated mice Kim, JW *et al.,* 2004, *Gene Ther.* 11:1011-1018. Thus, DNA vaccines employing CRT in this manner have the ability to enhance antigen-specific immune responses (as was originally demonstrated with the HPV E7 oncoprotein (see above).

<u>Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV).</u>

The present invention is directed to compositions and methods for stimulating immunity specific for the coronavirus responsible for severe acute respiratory syndrome (SARS). Eradication of SARS has become a priority for healthcare agencies around the world because of its communicability, associated mortality, and the potential for pandemic spread. As of July 31, 2003, 8098 cases had been identified worldwide and 774 had died, a mortality rate of about 9.6% (WHO statistics appear on the Web (at the URL who.int/csr/sars/country/table2003_09_23/en/) ; SARS has been attributed to infection with a coronavirus (SARS-CoV) (Drosten, C *et al.*, 2003, *N Engl J Med 348*:1967-76; Ksiazek, TG *et al.*, 2003, *N Engl J Med 348*:1953-66; Peiris, JS *et al.*, 2003, *Lancet 361*:1319-1225). Evidence that SARS-CoV is the etiologic agent of SARS was demonstrated by experimental infection of macaques (*Macaca fascicularis*), fulfilling Koch's postulates (Fouchier, RA, 2003. *Nature 423*:240). Knowledge of the structure of SARS-CoV and characterization of its complete RNA genome (Marra, MA *et al.*, 2003, *Science 300*:1399-404; Rota, PA *et al.*, 2003, *Science 300*:1394-1399; Ruan, YJ *et al.*, 2003, *Lancet 361*:1779-

1785) have provided the basic information that enabled the present inventors to develop[novel strategies for the prevention of SARS using vaccines.

Like its coronavirus relatives, SARS-CoV is a (+)-stranded RNA virus with a ~30kb genome encoding replicase (*rep*) gene products and structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N). S protein is thought to be involved with receptor binding, E protein plays a role in viral assembly, M is important for virus budding, and N protein is associated with viral RNA packaging (reviewed in Holmes, KV, 2003, *J. Clin. Invest. 111*:1605-1609. Among these proteins, it was not evident *a priori* which contain useful SARS-CoV-specific T cell epitopes or epitopes for targeting by neutralizing or protective antibodies. N protein was shown to generate coronavirus-specific CD8+ T cells, albeit in coronaviruses that infect non-human species (*i.e.,* mouse hepatitis virus and infectious bronchitis virus) and have different tissue tropism (Bergmann, C *et al.*, 1993, *J Virol 67*:7041-7049; Boots, AM *et al.*, 1991, *Immunology 74*:8-13; Seo, SH *et al.*, 1997, *J Virol 71*:7889-7894; Stohlman, SA *et al.*, 1992, *Virology 189*:217-224; Stohlman, SA *et al.*, 1993, *J Virol 67*:7050-7059). N-specific CD8+ T cells were shown to generate protective effects in other coronaviral systems (Collisson, EW *et al.*, 2000, *Dev Comp Immunol 24*:187-200; Seo *et al.*, *supra*).

SARS-CoV, spike (S) protein has been found to bind to angiotensin-converting enzyme 2 (ACE2), the functional receptor of SARS CoV on susceptible cells (Dimitrov, DS, 2003 *Cell 115*:652-653; Li, W *et al.*, 2003, *Nature 426*:450-454 ; Prabakaran, P *et al.*, 2004, *Biochem Biophys Res Commun. 314*:235-241; Wang, P *et al.*, 2004, *Biochem Biophys Res Commun. 315*:439-444). Analysis of the S protein has identified the receptor-binding domain, S1 (aa 1-680), and the membrane fusion domain, S2 (aa 680-1225) (see Figure 6) and SEQ ID NO:14-17. The receptor-binding domain S1 is responsible for binding to the ACE2 receptor (Dimitrov, *supra*; Li *et al., supra*; Prabakaran *et al., supra*; Wang *et al., supra*). Thus, innovative approaches interfering with the binding of S1 to ACE2, such as the immunological approaches disclosed herein, may protect the host from SARS CoV infection.

As a main surface antigen of SARS-CoV, was said to be one of the most important antigen candidates for vaccine design ((Zhao P *et al.*, 2004, *Acta Biochim Biophys Sin* (Shanghai) *36*:37-41). Vaccine strategies targeting the S protein of SARS-CoV have been developed. For instance, a highly attenuated modified vaccinia virus Ankara (MVA) has been engineered to express the S protein of SARS-CoV. Mice vaccinated with MVA-expressing S protein were capable of generating neutralizing antibodies (Bisht, H *et al.*, 2004, *Proc Natl Acad*

Sci USA 101:6641-6). In addition, a recombinant attenuated parainfluenza virus encoding SARS-CoV S protein has been shown to generate protective neutralizing antibodies in vaccinated mice (Buchholz, UJ *et al.*, 2004, *Proc Natl Acad Sci USA 101*:9804-98) and African green monkeys (Bukreyev, A, 2004, *Lancet 363*:2122-2127). Furthermore, a naked DNA vaccine encoding S protein generated protective neutralizing antibodies in vaccinated mice (Zhao *et al.*, *supra*). Three fragments of the truncated S protein were expressed in *E. coli* , and analyzed with pooled sera of convalescence phase of SARS patients. The full length S gene DNA vaccine was constructed and used to immunize BALB/c mice. The mouse serum IgG antibody against SARS-CoV was measured by ELISA with *E. coli* expressed truncated S protein or SARS-CoV lysate as diagnostic antigen. The results showed that all the three fragments of S protein expressed by *E. coli* were able to react with sera of SARS patients and the S gene DNA candidate vaccine could induce the production of specific IgG antibody against SARS-CoV efficiently in mice with seroconversion ratio of 75% after 3 times of immunization. As indicated elsewhere, while naked DNA vaccines in general have the clear advantages of simplicity, stability and safety over viral or bacterial vectors, they suffer from lack of potency, since they do not have the intrinsic ability to amplify and spread as live viral vectors do.

The present invention is focused on improved DNA vaccines comprising epitopes of any one or more of the S, E, M and N proteins of SARS-CoV.

SUMMARY OF THE INVENTION

The invention provides a nucleic acid encoding a chimeric protein comprising a first polypeptide domain comprising an endoplasmic reticulum chaperone polypeptide and a second polypeptide domain comprising at least one antigenic peptide. The antigenic peptide can comprise an MHC Class I-binding peptide epitope. The antigenic peptide, *e.g.*, the MHC class I-binding peptide epitope, can be between about 8 amino acid residues and about 11 amino acid residues in length.

The endoplasmic reticulum chaperone polypeptide includes any ER polypeptide having chaperone functions similar to the exemplary chaperones calreticulin, calnexin, tapasin, or ER60 polypeptides; or, analogues or mimetics thereof, or, functional fragments thereof. Such functional fragments can be screened using routine screening tests, *e.g.*, as described in Examples 1 and 2, below. Thus, in alternative embodiments, the endoplasmic reticulum chaperone polypeptide comprises or consists of a calnexin polypeptide or an equivalent thereof,

an ER60 polypeptide or an equivalent thereof, a GRP94/GP96 or a GRP94 polypeptide or an equivalent thereof, or, a tapasin polypeptide or an equivalent thereof.

In one embodiment, the calreticulin polypeptide comprises a human calreticulin polypeptide. In alternative embodiments, the human calreticulin polypeptide sequence can comprise SEQ ID NO:1, or, it can consist essentially of a sequence from about residue 1 to about residue 180 of SEQ ID NO:1, or, it can consist essentially of a sequence from about residue 181 to about residue 417 of SEQ ID NO:1.

In one embodiment, the isolated or recombinant nucleic acid molecule is operatively linked to a promoter, such as, *e.g.*, a constitutive, an inducible or a tissue-specific promoter. The promoter can be expressed in any cell, including cells of the immune system, including, *e.g.*, antigen presenting cells (APCs), *e.g.*, in a constitutive, an inducible or a tissue-specific manner.

In alternative embodiments, the APCs are dendritic cells, keratinocytes, astrocytes, monocytes, macrophages, B lymphocytes, a microglial cell, or activated endothelial cells, and the like.

The invention also provides an expression cassette comprising a nucleic acid sequence encoding a chimeric protein comprising a first polypeptide domain comprising an endoplasmic reticulum chaperone polypeptide and a second polypeptide domain comprising at least one antigenic peptide from a SARS-CoV. In alternative embodiments, the first domain comprises a calreticulin polypeptide and the second domain comprises an MHC class I-binding peptide epitope of a SARS-CoV antigen. In alternative embodiments, the expression cassette comprises an expression vector, a recombinant virus (*e.g.*, an adenovirus, a retrovirus), a plasmid. The expression cassette can comprise a self-replicating RNA replicon. The self-replicating RNA replicon can comprise a Sindbis virus self-replicating RNA vector, such as, *e.g.*, a Sindbis virus self-replicating RNA vector SINrep5 (U.S. Patent No. 5,217,879). As with all applicable embodiments of the invention, the ER chaperone polypeptide can include any ER polypeptide having chaperone functions similar to the exemplary chaperones calreticulin, 1, tapasin, or ER60 polypeptides; or, analogues or mimetics thereof, or, functional fragments thereof.

The invention also provides a particle comprising a nucleic acid encoding a chimeric protein comprising a first polypeptide domain comprising an endoplasmic reticulum chaperone polypeptide and a second polypeptide domain comprising at least one antigenic peptide. In one embodiment, the isolated particle comprising an expression cassette comprising a nucleic acid sequence encoding a fusion protein comprising at least two domains, wherein the first domain

comprises a calreticulin polypeptide and the second domain comprises an MHC class I-binding peptide epitope. The isolated particle can comprise any material suitable for particle bombardment, such as, *e.g.*, gold. The ER chaperone polypeptide can include any ER polypeptide having chaperone functions similar to the exemplary chaperones calreticulin, calnexin, tapasin, or ER60 polypeptides, as discussed herein.

The invention also provides a cell comprising a nucleic acid sequence encoding a chimeric protein comprising a first polypeptide domain comprising an endoplasmic reticulum chaperone polypeptide and a second polypeptide domain comprising at least one antigenic peptide. In one embodiment, the cell comprises an expression cassette comprising a nucleic acid sequence encoding a fusion protein comprising at least two domains, wherein the first domain comprises a calreticulin polypeptide and the second domain comprises an MHC class I-binding peptide epitope. The cell can be transfected, infected, transduced, *etc.*, with a nucleic acid of the invention or infected with a recombinant virus of the invention. The cell can be isolated from a non-human transgenic animal comprising cells comprising expression cassettes of the invention. Any cell can comprise an expression cassette of the invention, such as, *e.g.*, cells of the immune system or antigen presenting cells (APCs). The APCs can be a dendritic cell, a keratinocyte, a macrophage, a monocyte, a B lymphocyte, an astrocyte, a microglial cell, or an activated endothelial cell.

The invention also provides a chimeric polypeptide comprising a first polypeptide domain comprising an endoplasmic reticulum chaperone polypeptide, preferably human CRT, and a second polypeptide domain comprising at least one antigenic peptide of SARS-CoV. The antigenic peptide can comprise an MHC Class I-binding peptide epitope. The ER chaperone polypeptide can be chemically linked to the antigenic peptide, *e.g.*, as a fusion protein (*e.g.*, a peptide bond), that can be, *e.g.*, synthetic or recombinantly produced, *in vivo* or *in vitro*. The polypeptide domains can be linked by a flexible chemical linker.

In alternative embodiments, the first polypeptide domain of the chimeric polypeptide can be closer to the amino terminus than the second polypeptide domain, or, the second polypeptide domain can be closer to the amino terminus than the first polypeptide domain. The ER chaperone polypeptide can include any ER polypeptide having chaperone functions similar to the exemplary chaperones calreticulin, calnexin, tapasin, or ER60 polypeptides, as discussed herein.

The invention provides a pharmaceutical composition comprising a composition of the invention capable of inducing or enhancing an antigen specific immune response and a pharmaceutically acceptable excipient. In alternative embodiments, the composition comprises: a chimeric polypeptide comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising an antigenic peptide; a nucleic acid molecule encoding a fusion protein comprising a first polypeptide domain comprising an endoplasmic reticulum chaperone polypeptide and a second polypeptide domain an antigenic peptide; an expression cassette comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising an antigenic peptide; a particle comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising an antigenic peptide; or, a cell comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide coding sequence and a second domain comprising an antigenic peptide. The ER chaperone polypeptide can include any ER polypeptide having chaperone functions similar to the exemplary chaperones calreticulin, calnexin, tapasin, or ER60 polypeptides, as discussed herein.

The invention provides a method of inducing or enhancing an antigen specific immune response comprising: (a) providing a composition comprising a composition of the invention capable of inducing or enhancing an antigen specific immune response, which, in alternative embodiments, can be: a chimeric polypeptide comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising an antigenic peptide; a nucleic acid molecule encoding a fusion protein comprising a first polypeptide domain comprising an endoplasmic reticulum chaperone polypeptide and a second polypeptide domain an antigenic peptide; an expression cassette comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising an antigenic peptide; a particle comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising an antigenic peptide; or, a cell comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide coding sequence and a second domain comprising an antigenic peptide; and, (b) administering an

amount of the composition sufficient to induce or enhance an antigen specific immune response. The antigen specific immune response can comprise cellular response, such as a $CD8^+$ CTL response. The antigen specific immune response can also comprise an antibody-mediated response, or, a humoral and a cellular response.

In practicing the method the composition can administered *ex vivo,* or, the composition can be administered *ex vivo* to an antigen presenting cell (APC). In alternative embodiments, the APC is a dendritic cell, a keratinocyte, a macrophage, a monocyte, a B lymphocyte, an astrocyte, a microglial cell, or an activated endothelial cell. The APC can be a human cell. The APC can be isolated from an *in vivo* or *in vitro* source. The method can further comprise administering the *ex vivo*-treated APC to a mammal, a human, a histocompatible individual, or to the same individual from which it was isolated. Alternatively, the composition is administered directly *in vivo* to a mammal, *e.g.*, a human.

The composition can be administered intramuscularly, intradermally, or subcutaneously. The composition ,*e.g.*, the nucleic acid, expression cassette or particle, can be administered by biolistic injection.

The invention provides a method of increasing the numbers of $CD8^+$ CTLs specific for a desired SARS-CoV antigen in an individual comprising: (a) providing a composition comprising: a chimeric polypeptide comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide, preferably CRT, and a second domain comprising an antigenic peptide of SARS-CoV; a nucleic acid molecule encoding a fusion protein comprising a first polypeptide domain comprising an endoplasmic reticulum chaperone polypeptide and a second polypeptide domain the antigenic peptide; an expression cassette comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising the antigenic peptide; a particle comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising the antigenic peptide; or, a cell comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide coding sequence and a second domain comprising the antigenic peptide; wherein the MHC class I-binding peptide epitope is derived from a SARS-CoV antigen, preferably the S protein, the M protein, the N protein or the E protein , and, (b) administering an amount of the composition sufficient to increase the numbers of antigen-specific $CD8^+$ CTL.

The invention provides a method of inhibiting a SARS-CoV infection or spread of the virus in a subject comprising: (a) providing a composition comprising: a chimeric polypeptide comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising a SAR-CoV antigenic peptide; a nucleic acid molecule encoding a fusion protein comprising a first polypeptide domain comprising an endoplasmic reticulum chaperone polypeptide and a second polypeptide domain the antigenic peptide; an expression cassette comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising the antigenic peptide; a particle comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide and a second domain comprising the antigenic peptide; or, a cell comprising a nucleic acid sequence encoding a fusion protein comprising a first domain comprising an endoplasmic reticulum chaperone polypeptide coding sequence and a second domain comprising the antigenic peptide; and, (b) administering an amount of the composition sufficient to inhibit the infection or spread of the virus in vivo. The composition can be co-administered with a second composition that has antiviral activity.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

All publications, patents, patent applications, GenBank sequences and ATCC deposits, cited herein are hereby expressly incorporated by reference for all purposes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a Western blot that characterizes recombinant SARS-CoV N protein expression in 293 cells transfected with pcDNA3.1/myc-His (-) encoding CRT, N, CRT/N, or no insert. Rabbit anti-GST-N sera was used at a 1:100 dilution to detect N expression. Lane 1: lysate from 293 cells transfected with pcDNA3.1/myc-His (-); Lane 2: lysate from 293 cells transfected with CRT DNA; Lane 3: lysate from 293 cells transfected with N DNA; Lane 4: lysate from 293 cells transfected with CRT/N DNA.

Figures 2A-2D are a gel, a blot and graphs showing the N-specific humoral immune response in mice vaccinated with various nucleic acid preparations. Fig. 2A shows a Coomassie blue-stained SDS-PAGE gel of N protein purified from *E. coli*. Lane 1: marker; Lane

2: crude extract of *E. coli* expressing N protein; Lane 3: purified GST-N protein. Fig. 2B shows a Western blot confirming the presence of purified GST-N protein. Lane 1: lysate from 293 cells transfected with plasmid DNA without an insert (negative control) Lane 2: lysate from 293 cells transfected with plasmid DNA encoding N protein (positive control) Lane 3: purified GST-N protein. Fig. 2C shows results of ELISA determining the titers of N-specific IgG antibodies in sera from vaccinated mice. Sera were collected from DNA-vaccinated mice (5/group) one week after the last vaccination and antibodies against bacteria-derived GST N protein were tested. Purified GST protein was used as a control. Sera from vaccinated mice only generated background level of color changes against GST (not shown). Fig. 2D shows results of an ELISA comparing the relative titers of N-specific IgG1 and IgG2a antibodies in sera of DNA-vaccinated mice (5/group).

Figures 3A-3C are flow cytometric tracings and graphs showing SARS-CoV N-specific CD8+ T cell mediated immune responses in mice vaccinated with the various DNA compositions. Intracellular cytokine staining followed by flow cytometry analysis was used to characterize the N-specific $CD8^+$ T cell response to vaccination. **Fig. 3A** shows a representative flow cytometric analysis. Fig. 3B depicts the number of SARS-CoV N peptide-specific IFN-γ-secreting CD8+ T cell precursors (per $3x10^5$ splenocytes) stimulated by the indicated peptide *in vitro* after harvesting from spleens of mice vaccinated with CRT/N DNA (5 per group). The peptides derived from SARS-CoV N protein are defined in Table 3. Fig. 3C is a graph depicting the number of N-specific IFN-γ-secreting CD8+ T cell precursors/$3x10^5$ splenocytes in spleen cells harvested from mice (5 per group) that had been vaccinated with various DNA constructs as indicated: plasmid DNA encoding N, CRT, CRT/N or lacking any insert were cultured with MHC class I-restricted N peptide (aa 346-354, QFKDNVILL (SEQ ID NO:31 *in vitro* overnight and stained for CD8 and IFN-γ.

Figures 4A-4C shows SARS-CoV N protein expression in cells infected with recombinant N vaccinia . 293 cells were infected with either wild type vaccinia virus (Vac-WT) or vaccinia virus expressing SARS N protein (Vac-N). Rabbit anti-GST-N sera was used to identify N protein expression. Fig.43A shows a flow cytometric analysis. Fig.4B shows immunofluorescence staining. Fig. 4C shows a Western blot using cell lysate from 293 cells infected with either Vac-WT (**Lane 1**)or Vac-N (**Lane 2**). Note: Lysate from 293 cells infected with Vac-N revealed a band approximately M_r 48,000 in size, corresponding to N protein of SARS-CoV.

Figures 5A-5B are graphs showing reduction of the viral titer of recombinant N vaccinia in mice vaccinated with the various DNA vaccines. Mice (5 per group) were vaccinated with pcDNA3.1/myc-His (-) encoding CRT, N, CRT/N, or no insert as described in the Examples. Fig. 1 A shows virus titers after intranasal challenge with vaccinia. The immunized mice were infected with 2×10^6 PFU/mouse of Vac-WT or Vac-N in 20 μl by intranasal instillation 1 week after the final immunization. Vac-WT infection was used as a negative control. Fig. 5B shows results of i.v. challenge with vaccinia. The immunized mice were infected with 10^7 PFU/mouse of Vac-N in 100 μl by intravenous injection 1 week after final immunization. The titer of virus was determined by plaque assay 5 days after challenge. Note: Mice vaccinated with CRT/N DNA showed the greatest reduction in titer of Vac-N virus when challenged intranasally or intravenously.

Figure 6 is a schematic diagram of SARS-CoV S protein showing a domain structure. Domain S1 corresponds to residues 1-680 of SEQ ID NO:14; with residues 1-18 representing a signal sequence), S2 corresponds to residues 681-1225 of SEQ ID NO:14 and includes two helical regions (HR1 and HR2) as well as a transmembrane domain. Si represents an overlapping fragment of S1 and S2, and includes residues 417-816 or SEQ ID NO:14; (polypeptide indicating and its recombinants used for immunization. Recombinant nucleic acids comprising S1, S2 and Si were examined as immunogens.

Figure 7A-7B show blots that represents expression and secretion of SARS-CoV S and its recombinant proteins after *in vitro* transfection . The expression of SARS-CoV S and its recombinant proteins was determined in 293 cells transfected with a DNA molecule encoding S, S1, Si or S2 by Western blot analysis (Fig. 7A). Overnight after transfection, the cells were lysed with protein extraction reagent (Pierce, Rockford, IL). Equal amounts of proteins (50 μg) were loaded and separated by 10% SDS-PAGE. Rabbit anti-S antibody at a 1:2000 dilution was used to detect expression of the full length S polypeptide and its recombinant domains/fragments. The presence of secreted SARS-CoV S proteins and recombinant domains confirmed by Western blot analysis (Fig 7B). Forty eight hours after transfection, 4 ml of culture supernatants were collected, centrifuged to remove cellular debris and concentrated to 0.2 ml using Amicon Ultra centrifugal filter devices. Concentrated supernatants (20 μl) were loaded and separated by 10% SDS-PAGE before blotting. The presence of S and its recombinant domains/fragments proteins was detected as above.

Figure 8A-8B shows results of an S-specific antibody responses in mice immunized with various recombinant SARS-CoV S DNA immunogens. Mice were immunized with the plasmid DNAs encoding S, S1, Si or S2 via gene gun. Serum samples were collected from one week after the last vaccination and tested for anti-S antibodies. S-specific antibodies were detected in serum diluted to 1:250 (in PBS) by Western blot analysis using 50 μg of transfected 293 lysates with DNA encoding S (**Fig. 8A**). The end-point dilution titer of S-specific antibodies in the sera of DNA-immunized C57BL/6 mice were determined by ELISA in microplates coated with "TC-1/S" cells or "TC-1/No insert" cells (**Fig. 8B**). **Absorbances** >3-fold higher than negative controls were considered positive.

Figure 9A-9B show SARS-CoV S-specific CD8+ T cell responses in mice immunized with the various DNA immunogens. Intracellular cytokine staining (IFNγ = INFγ) was determined after flow cytometry to characterize the S-specific $CD8^+$ T cell response. Fig. 9**A** shows flow cytometric analysis and **Fig. 9B** is a bar graph depicting the number of IFNγ-secreting $CD8^+$ T cell precursors /3×10^5 splenocytes. $CD3^+$ cells (10^6) were harvested from spleens of immunized given S, S1, Si or S2-encoding DNA immunogens. These cells were stimulated with 10^5 "DC/S" dendritic cells or "DC/No insert" dendritic cells *in vitro* overnight and were stained for CD8 and IFNγ as measures of SARS-CoV S-specific $CD8^+$ T cell immunity.

Figure 10A-10B show expression and secretion of S1 and CRT/S1 chimeric polypeptide after *in vitro* transfection. Expression was determined in 293 cells transfected with DNA constructs comprising no insert, CRT, S1 or CRT/S1 by Western blot analysis (**Fig. 10A**). After overnight incubation, transfected cells were lysed and equal amounts of proteins (50 μg) were loaded and separated by 10% SDS-PAGE. Rabbit anti-S antibody diluted 1:2000 was used to detect S1 and the CRT/S1 chimeric polypeptide. The presence of secreted S1 and CRT/S1 was also examined by Western blot analysis (**Fig. 10B**). Forty eight hours after transfection, 4 ml of culture supernatants were obtained, centrifuged and concentrated as above. Samples (5, 10, 20 μl) of the concentrated supernatants were separated by SDS-10% PAGE before blotting. Detection was as above with rabbit anti-S antibody.

Figure 11A-11B shows that immunization with DNA encoding CRT/S1 induces a stronger antibody responses than DNA encoding alone. Mice were immunized with the plasmid DNAs encoding no insert, CRT, S1 or CRT/S1 via gene gun. Serum samples were collected and antibodies measure as described for Fig. 8A-8AB.

Figure 12A-12B shows that more potent SARS-CoV S-specific CD8+ T cell responses result from administration of DNA immunogens encoding the CRT/S1 fusion protein. Methods are the same as described for Fig. 9A-9B.

Figure 13A-13B shows that mice vaccinated with DNA immunogens encoding the chimeric polypeptide CRT/S1 have stronger *in vivo* protection against growth of a tumor expressing the SARS-CoV S protein. **Fig. 13A** shows a study in which transfected tumor cells expressing S (TC-1/S) were injected subcutaneously (5×10^5cells/mouse) into mice that had been immunized with a DNA constructs that encoded CRT, S1, CRT/S1 or no insert (10 mice/group). Animals received the challenge in the right leg one week after the last vaccination and were monitored twice weekly for visible tumor. **Fig 13B** shows results of tumor growth when various subsets of immune cells were depleted by antibody treatment *in vivo*. CD4, CD8, and NK1.1 depletion was initiated one week after last vaccination and the mice challenged one week later. The depletion treatment was terminated 32 days after tumor challenge. For each time point shown, >99% of the appropriate cell subset was depleted with normal numbers of cells of other subsets.

Figure 14. is a Western blot that characterizes recombinant SARS-CoV M (membrane) protein expression in 293 cells transfected with pcDNA3.1/myc-His (-) encoding CRT, M or CRT/M. pcDNA3.1/myc-His (-) without insert was used as a negative control. The transfected cells were lysed 24 hours later and separated by SDS-PAGE. Mouse anti-myc antibody was used to detect M protein expression. Lanes 1-4 show lysates from 293 cells transfected with DNA without an insert and DNA encoding CRT, M or CRT/M, respectively.

Figure 15A-15B show SARS-CoV M-specific CD8+ T cell responses in mice immunized with the various DNA immunogens encoding the M polypeptide. Five mice per group were immunized with pcDNA3, pcDNA3-CRT, pcDNA3-M or pcDNA3-CRT/M. $CD3^+$ enriched T cells from spleens of immunized mice were stimulated *in vitro* with transfected dendritic cells, DC/S" dendritic cells or "DC/No insert", *in vitro* overnight and stained for both CD8 and intracellular IFNγ. **Fig. 15A** shows representative flow cytometry results for $CD3^+$ enriched T cells from immunized or control mice. **Fig. 15B is a b**ar graph depicting the number of antigen-specific IFNγ-secreting $CD8^+$ T-cell precursors/3×10^5 $CD3^+$ enriched T cells (mean±SD) after DNA vaccination.

Figure 16A-16B presents flow cytometric analysis of IFN-γ-secreting M-specific $CD4^+$ T-cells (Th1) in mice (five per group) immunized with pcDNA3, pcDNA3-CRT, pcDNA3-M or

pcDNA3-CRT/M. $CD3^+$ -enriched T cells from spleens of immunized mice were stimulated *in vitro* with DC-1/M or DC-1/no insert overnight, and stained for both CD4 and intracellular IFNγ. **Fig. 16A** presents representative flow cytometry data for splenocytes harvested from immunized mice. **Fig. 16B** is a bar graph depicting the number of antigen-specific IFNγ-secreting $CD4^+$ T-cells (Th1 cells) per $3x10^5$ $CD3^+$ enriched T cells (mean±SD).

Figure 17A-17B presents flow cytometry analysis of IL-4-secreting M-specific $CD4^+$ T-cells (Th2) in mice (five per group) immunized with pcDNA3, pcDNA3-CRT, pcDNA3-M or pcDNA3-CRT/M. CD3+ enriched T cells from spleens of immunized mice were stimulated *in vitro* with DC-1/M or DC-1/no insert overnight, and stained for both CD4 and intracellular IL-4. **Fig. 17A** presents representative flow cytometry data for splenocytes harvested from immunized mice. **Fig. 17B** presents a bar graph depicting the number of antigen-specific IL-4-secreting $CD4^+$ T-cells (Th2 cells) per $3x10^5$ CD3+ enriched T cells (mean±SD).

Figure 18A-18B shows that mice vaccinated with DNA immunogens encoding the chimeric polypeptide CRT/M are much better protected *in vivo* against growth of a tumor expressing the SARS-CoV M protein. **Fig. 18A** shows a study in which transfected tumor cells expressing M (TC-1/M) were injected subcutaneously (5×10^4cells/mouse) into mice that had been immunized with a plasmid DNA constructs that encoded (i) CRT, (ii) M, (iii) CRT/M or (iv) no insert (10 mice/group). Animals received the challenge in the right leg one week after the last vaccination and were monitored twice weekly for visible tumor. **Fig 18B** shows results of tumor growth when various subsets of immune cells were depleted by antibody treatment *in vivo*. CD4, CD8, and NK1.1 depletion was initiated one week after last vaccination and the mice challenged one week later. The depletion treatment was terminated 32 days after tumor challenge. Both graphs show the percentage of tumor-free mice over time.

Figure 19 shows schematically **SARS-CoV cDNA clones spanning the genome of the TW1 strain.**

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The invention provides compositions and methods for enhancing the immune responses, particularly cytotoxic T cell immune responses, induced *in vivo* administration of chimeric nucleic acids that encode (a) an endoplasmic reticulum chaperone polypeptide linked to (b) at least one antigenic polypeptide or peptide from SARS CoV. These chimeric polypeptides or

fusion proteins can also be administered, although the preferred embodiment is a nucleic acid composition or expression plasmid for administration as an immunogen or vaccine.

For descriptions of this general strategy as using chaperone polypeptides or other such polypeptides to enhance the potency of a vector carrying antigen-encoding DNA, see for example, Wu *et al.*, WO 01/29233; Wu *et al.*, WO 02/009645; Wu *et al.*, WO 02/061113; Wu *et al.*, WO 02/074920; Wu *et al.*, WO 02/12281, all of which are incorporated by reference in their entirety.

The fusion polypeptide encoded by the nucleic acid immunogenic or vaccine composition comprises at least two "domains:" the first domain comprises a endoplasmic reticulum chaperone polypeptide, and the second domain comprises a full length polypeptide or a shorter fragment that comprises at least one epitope-comprising a SARS-CoV structural protein, most preferably the product of the S, E, M or N gene of SARS-CoV.

Although any endoplasmic reticulum chaperone polypeptide, or functional fragment or variation thereof, can be used in the invention, such as calreticulin, tapasin, ER60 or calnexin polypeptides, human calreticulin (CRT) is preferred.

The antigenic domain of the chimeric molecule is preferably one that comprises an MHC class I-binding peptide epitope.

In the methods of the invention, the chimeric nucleic acid or polypeptide are administered or applied to induce or enhance immune responses that are specific and anti-viral in their effect (e.g., that neutralize virus or result in damage and death of virus expressing cells) *in vivo*.

The experiments described herein demonstrate that the methods of the invention can enhance a cellular immune response, particularly, a CTL reactivity, induced by a DNA vaccine encoding various polypeptides of the SARS CoV. Initially, DNA encoding the nucleocapsid or N-protein was used. .

As described in Example 1, below, the results of these experiments demonstrate that DNA vaccines comprising nucleic acid encoding a fusion protein comprising CRT linked to a N protein of SARS-CoV enhances the potency of DNA vaccines. DNA vaccines of the invention containing chimeric CRT fusion genes were or will be administered to mice and other subjects by biolistic subcutaneous methods. They induced increased N-specific CD8+ CTL precursors, and are expected to improve immune protection against the virus. This increase in N-specific

CD8+ T cell precursors was significant as compared to DNA vaccines containing N or CRT genes alone.

A potential mechanism for the enhanced antigen-specific $CD8^+$ T cell immune responses *in vivo* is the presentation of antigen through the MHC class I pathway by uptake of apoptotic bodies from cells expressing the antigen, also called "cross-priming".

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

The term "antigen" or "immunogen" as used herein refers to a compound or composition comprising a peptide, polypeptide or protein which is "antigenic" or "immunogenic" when administered (or expressed *in vivo* by an administered nucleic acid, *e.g.*, a DNA vaccine) in an appropriate amount (an "immunogenically effective amount"), *i.e.*, is capable of eliciting, augmenting or boosting a cellular and/or humoral immune response either alone or in combination or linked or fused to another substance (which can be administered at once or over several intervals).

"Calnexin" describes the well-characterized membrane protein of the endoplasmic reticulum (ER) that functions as a molecular chaperone and as a component of the ER quality control machinery. Calreticulin is a soluble analogue of calnexin. *In vivo,* calreticulin and calnexin play important roles in quality control during protein synthesis, folding, and posttranslational modification. Calnexin polypeptides, and equivalents and analogues thereof, are species in the genus of ER chaperone polypeptides, as described herein (Wilson (2000) J. Biol. Chem. 275:21224-2132; Danilczyk (2000) J. Biol. Chem. 275:13089-13097; U.S. Patent Nos. 6,071,743 and 5,691,306).

"Calreticulin" or "CRT" describes the well-characterized ~46 kDa resident protein of the ER lumen that has lectin activity and participates in the folding and assembly of nascent glycoproteins. CRT acts as a "chaperone" polypeptide and a member of the MHC class I transporter TAP complex; CRT associates with TAP1 and TAP2 transporters, tapasin, MHC Class I heavy chain polypeptide and β2 microglobulin to function in the loading of peptide epitopes onto nascent MHC class I molecules (Jorgensen (2000) Eur. J. Biochem. 267:2945-2954). The term "calreticulin" or "CRT" refers to polypeptides and nucleic acids molecules

having substantial identity (defined herein) to the exemplary CRT sequences as described herein. A CRT polypeptide is a polypeptides comprising a sequence identical to or substantially identical (defined herein) to the amino acid sequence of CRT. An exemplary nucleotide and amino acid sequence for a CRT used in the present compositions and methods are SEQ ID NO:1 and SEQ ID NO:2, respectively. The terms "calreticulin" or "CRT" encompass native proteins as well as recombinantly produced modified proteins that induce an immune response, including a CTL response. The terms "calreticulin" or "CRT" encompass homologues and allelic variants of CRT, including variants of native proteins constructed by *in vitro* techniques, and proteins isolated from natural sources. The CRT polypeptides of the invention, and sequences encoding them, also include fusion proteins comprising non-CRT sequences, particularly MHC class I-binding peptides; and also further comprising other domains, *e.g.*, epitope tags, enzyme cleavage recognition sequences, signal sequences, secretion signals and the like.

The term "endoplasmic reticulum chaperone polypeptide" as used herein means any polypeptide having substantially the same ER chaperone function as the exemplary chaperone proteins CRT, tapasin, ER60 or calnexin. Thus, the term includes all functional fragments or variants or mimics thereof. A polypeptide or peptide can be routinely screened for its activity as an ER chaperone using assays known in the art. While the invention is not limited by any particular mechanism of action, *in vivo* chaperones promote the correct folding and oligomerization of many glycoproteins in the ER, including the assembly of the MHC class I heterotrimeric molecule (heavy chain, β2m, and peptide). They also retain assembled MHC class I heterotrimeric complexes in the ER (Hauri (2000) FEBS Lett. 476:32-37).

The term "epitope" as used herein refers to an antigenic determinant or antigenic site that interacts with an antibody or a T cell receptor (TCR), *e.g.*, the MHC class I-binding peptide compositions used in the methods of the invention. An "antigen" is a molecule or chemical structure that either induces an immune response or is specifically recognized or bound by the product of an immune response, such as an antibody or a CTL. The specific conformational or stereochemical "domain" to which an antibody or a TCR bind is an "antigenic determinant" or "epitope." TCRs bind to peptide epitopes which are physically associated with a third molecule, a major histocompatibility complex (MHC) class I or class II protein.

The terms "ER60" or "GRP94" or "gp96" or "glucose regulated protein 94" as used herein describes the well-characterized ER chaperone polypeptide that is the ER representative of the heat shock protein-90 (HSP90) family of stress-induced proteins. These bind to a limited

number of proteins in the secretory pathway, possibly by recognizing advanced folding intermediates or incompletely assembled proteins. ER60 polypeptides, and equivalents and analogues thereof, are species in the genus of ER chaperone polypeptides, as described herein (Argon (1999) *Semin. Cell Dev. Biol.* 10:495-505; Sastry (1999) J. Biol. Chem. *274*:12023-12035; Nicchitta (1998) *Curr. Opin. Immunol.* 10:103-109; U.S. Patent No. 5,981,706).

The term "expression cassette" or "expression vector" as used herein refers to a nucleotide sequence which is capable of affecting expression of a protein coding sequence in a host compatible with such sequences. Expression cassettes include at least a promoter operably linked with the polypeptide coding sequence; and, optionally, with other sequences, *e.g.*, transcription termination signals. Additional factors necessary or helpful in effecting expression may also be included, *e.g.*, enhancers. "Operably linked" refers to linkage of a promoter upstream from a DNA sequence such that the promoter mediates transcription of the DNA sequence. Thus, expression cassettes include plasmids, recombinant viruses, any form of a recombinant "naked DNA" vector, and the like. A "vector" comprises a nucleic acid which can infect, transfect, transiently or permanently transduce a cell. It will be recognized that a vector can be a naked nucleic acid, or a nucleic acid complexed with protein or lipid. The vector optionally comprises viral or bacterial nucleic acids and/or proteins, and/or membranes (*e.g.*, a cell membrane, a viral lipid envelope, *etc.*).

Vectors include, but are not limited to replicons (*e.g.*, RNA replicons), bacteriophages) to which fragments of DNA may be attached and become replicated. Vectors thus include, but are not limited to RNA, autonomous self-replicating circular or linear DNA or RNA, *e.g.*, plasmids, viruses, and the like (U.S. Patent No. 5,217,879), and includes both the expression and nonexpression plasmids. Where a recombinant microorganism or cell culture is described as hosting an "expression vector" this includes both extrachromosomal circular and linear DNA and DNA that has been incorporated into the host chromosome(s). Where a vector is being maintained by a host cell, the vector may either be stably replicated by the cells during mitosis as an autonomous structure, or is incorporated within the host's genome.

The term "chemically linked" refers to any chemical bonding of two moieties, *e.g.*, as in one embodiment of the invention, where an ER chaperone polypeptide or CRT is chemically linked to an antigenic peptide. Such chemical linking includes the peptide bonds of a recombinantly or *in vivo* generated fusion protein.

The term "chimeric" or "fusion" polypeptide or protein refers to a composition comprising at least one polypeptide or peptide sequence or domain which is associated with a second polypeptide or peptide domain. One embodiment of this invention is an isolated or recombinant nucleic acid molecule encoding a fusion protein comprising at least two domains, wherein the first domain comprises an endoplasmic reticulum chaperone, *e.g.*, CRT, and the second domain comprising an antigenic epitope, *e.g.*, an MHC class I-binding peptide epitope. Additional domains can comprise a polypeptide, peptide, polysaccharide, or the like. The "fusion" can be an association generated by a peptide bond, a chemical linking, a charge interaction (*e.g.*, electrostatic attractions, such as salt bridges, H-bonding, *etc.*) or the like. If the polypeptides are recombinant, the "fusion protein" can be translated from a common message. Alternatively, the compositions of the domains can be linked by any chemical or electrostatic means. The chimeric molecules of the invention (*e.g.*, CRT-class I-binding peptide fusion proteins) can also include additional sequences, *e.g.*, linkers, epitope tags, enzyme cleavage recognition sequences, signal sequences, secretion signals, and the like. Alternatively, a peptide can be linked to a carrier simply to facilitate manipulation or identification/ location of the peptide.

The term "immunogen" or "immunogenic composition" refers to a compound or composition comprising a peptide, polypeptide or protein which is "immunogenic," *i.e.*, capable of eliciting, augmenting or boosting a cellular and/or humoral immune response, either alone or in combination or linked or fused to another substance. An immunogenic composition can be a peptide of at least about 5 amino acids, a peptide of 10 amino acids in length, a fragment 15 amino acids in length, a fragment 20 amino acids in length or greater; smaller immunogens may require presence of a "carrier" polypeptide *e.g.*, as a fusion protein, aggregate, conjugate or mixture, preferably linked (chemically or otherwise) to the immunogen. The immunogen can be recombinantly expressed from a vaccine vector, which can be naked DNA comprising the immunogen's coding sequence operably linked to a promoter, *e.g.*, an expression cassette. The immunogen includes one or more antigenic determinants or epitopes which may vary in size from about 3 to about 15 amino acids. Epitopes of more than one SARS-CoV protein may be used in combination.

The term "isolated" as used herein, when referring to a molecule or composition, such as, *e.g.*, a CRT nucleic acid or polypeptide, means that the molecule or composition is separated from at least one other compound, such as a protein, other nucleic acids (*e.g.*, RNAs), or other

contaminants with which it is associated *in vivo* or in its natural state. Thus, a CRT composition is considered isolated when it has been isolated from any other component with which it is natively associated, *e.g.*, cell membrane, as in a cell extract. An isolated composition can, however, also be substantially pure. An isolated composition can be in a homogeneous state and can be dry or in an aqueous solution. Purity and homogeneity can be determined, for example, using analytical chemistry techniques such as polyacrylamide gel electrophoresis (SDS-PAGE) or high performance liquid chromatography (HPLC). Thus, the isolated compositions of this invention do not contain materials normally associated with their *in situ* environment. Even where a protein has been isolated to a homogenous or dominant band, there are trace contaminants which co-purify with the desired protein.

The terms "polypeptide," "protein," and "peptide" include compositions of the invention that also include "analogues ," or "conservative variants" and "mimetics" or "peptidomimetics" with structures and activity that substantially correspond to the polypeptide from which the variant was derived, including, *e.g.*, human CRT or a Class I-binding peptide epitope, such as from the SARS-CoV S, E, M or N proteins. as discussed in detail, below.

The term "pharmaceutical composition" refers to a composition suitable for pharmaceutical use, *e.g.*, as a vaccine, in a subject. The pharmaceutical compositions of this invention are formulations that comprise a pharmacologically effective amount of a composition comprising, *e.g.*, a nucleic acid, or vector, or cell of the invention, and a pharmaceutically acceptable carrier.

The term "promoter" is an array of nucleic acid control sequences which direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter which is active under most environmental and developmental conditions. An "inducible" promoter is a promoter which is under environmental or developmental regulation. A "tissue specific" promoter is active in certain tissue types of an organism, but not in other tissue types from the same organism. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic

acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The term "recombinant" refers to (1) a polynucleotide synthesized or otherwise manipulated *in vitro* (*e.g.*, "recombinant polynucleotide"), (2) methods of using recombinant polynucleotides to produce gene products in cells or other biological systems, or (3) a polypeptide ("recombinant protein") encoded by a recombinant polynucleotide. For example, recombinant CRT or an MHC class I-binding peptide epitope can be recombinant as used to practice this invention. "Recombinant means" also encompass the ligation of nucleic acids having various coding regions or domains or promoter sequences from different sources into an expression cassette or vector for expression of, *e.g.*, inducible or constitutive expression of polypeptide coding sequences in the vectors used to practice this invention.

The term "self-replicating RNA replicon" refers to constructs based on RNA viruses, *e.g.*, alphavirus genome RNAs (*e.g.*, Sindbis virus, Semliki Forest virus, *etc.*), that have been engineered to allow expression of heterologous RNAs and proteins. These recombinant vectors are self-replicating (*i.e.*, they are "replicons") and can be introduced into cells as naked RNA or DNA, as described in detail, below. In one embodiment, the self-replicating RNA replicon comprises a Sindbis virus self-replicating RNA vector SINrep5, which is described in detail in U.S. Patent No. 5,217,879.

The term "systemic administration" refers to administration of a composition or agent such as the molecular vaccine or the CRT-Class I-binding peptide epitope fusion protein described herein, in a manner that results in the introduction of the composition into the subject's circulatory system. The term "regional" administration refers to administration of a composition into a specific anatomical space, such as intraperitoneal, intrathecal, subdural, or to a specific organ, and the like. For example, regional administration includes administration of the composition or drug into the hepatic artery. The term "local administration" refers to administration of a composition or drug into a limited, or circumscribed, anatomic space, such as intratumoral injection into a tumor mass, subcutaneous injections, intramuscular injections, and the like. Any one of skill in the art would understand that local administration or regional administration may also result in entry of the composition or drug into the circulatory system.

"Tapasin" is the known ER chaperone polypeptide, as discussed above. While not limited by any particular mechanism of action, *in vivo*, tapasin is a subunit of the TAP (transporter associated with antigen processing) complex and binds both to TAP1 and MHC

class I polypeptides. Tapasin polypeptides, and equivalents and analogues thereof, are species in the genus of ER chaperone polypeptides, as described herein (Barnden (2000) *J. Immunol.* 165:322-330; Li (2000) *J. Biol. Chem.* 275:1581-1586).

Generating and Manipulating Nucleic Acids

The methods of the invention provide for the administration of nucleic acids encoding a CRT-SARS-CoV Class I epitope binding peptide fusion protein, as described above. Recombinant CRT-containing fusion proteins can be synthesized *in vitro* or *in vivo*. Nucleic acids encoding these compositions can be in the form of "naked DNA" or they can be incorporated in plasmids, vectors, recombinant viruses (*e.g.*, "replicons") and the like for *in vivo* or *ex vivo* administration. Nucleic acids and vectors of the invention can be made and expressed *in vitro* or *in vivo*, a variety of means of making and expressing these genes and vectors can be used. One of skill will recognize that desired gene activity can be obtained by modulating the expression or activity of the genes and nucleic acids (*e.g.*, promoters) within vectors used to practice the invention. Any of the known methods described for increasing or decreasing expression or activity, or tissue specificity, of genes can be used for this invention. The invention can be practiced in conjunction with any method or protocol known in the art, which are well described in the scientific and patent literature.

General Techniques

The nucleic acid sequences used to practice this invention, whether RNA, cDNA, genomic DNA, vectors, recombinant viruses or hybrids thereof, may be isolated from a variety of sources, genetically engineered, amplified, and/or expressed recombinantly. Any recombinant expression system can be used, including, in addition to bacterial cells, *e.g.*, mammalian, yeast, insect or plant cell expression systems. Alternatively, these nucleic acids can be synthesized *in vitro* by well-known chemical synthesis techniques, as described in, *e.g.*, Carruthers (1982) *Cold Spring Harbor Symp. Quant. Biol. 47*:411-418; Adams (1983) *J. Am. Chem. Soc. 105*:661; Belousov (1997) *Nucleic Acids Res. 25*:3440-3444; Frenkel (1995) *Free Radic. Biol. Med. 19*:373-380; Blommers (1994) *Biochemistry 33*:7886-7896; Narang (1979) *Meth. Enzymol. 68*:90; Brown (1979) *Meth. Enzymol. 68:109*; Beaucage (1981) *Tetra. Lett. 22*:1859; U.S. Patent No. 4,458,066. Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Calreticulin Sequences

The sequences of CRT, including human CRT, are well known in the art (McCauliffe (1990) J. Clin. Invest. 86:332-335; Burns (1994) Nature 367:476-480; Coppolino (1998) Int. J. Biochem. Cell Biol. 30:553-558). The nucleic acid sequence appears as GenBank Accession No. NM 004343 and is SEQ ID NO:1.

```
1    gtccgtactg cagagccgct gccggagggt cgttttaaag ggccgcgttg ccgcccctc
61   ggcccgccat gctgctatcc gtgccgctgc tgctcggcct cctcggcctg gccgtcgccg
121  agcccgccgt ctacttcaag gagcagtttc tggacggaga cgggtggact tcccgctgga
181  tcgaatccaa acacaagtca gattttggca aattcgttct cagttccggc aagttctacg
241  gtgacgagga gaaagataaa ggtttgcaga caagccagga tgcacgcttt tatgctctgt
301  cggccagttt cgagcctttc agcaacaaag gccagacgct ggtggtgcag ttcacggtga
361  aacatgagca gaacatcgac tgtgggggcg gctatgtgaa gctgtttcct aatagtttgg
421  accagacaga catgcacgga gactcagaat acaacatcat gtttggtccc gacatctgtg
481  gccctggcac caagaaggtt catgtcatct tcaactacaa gggcaagaac gtgctgatca
541  acaaggacat ccgttgcaag gatgatgagt ttacacacct gtacacactg attgtgcggc
601  cagacaacac ctatgaggtg aagattgaca acagccaggt ggagtccggc tccttggaag
661  acgattggga cttcctgcca cccaagaaga taaaggatcc tgatgcttca aaaccggaag
721  actgggatga gcgggccaag atcgatgatc ccacagactc caagcctgag gactgggaca
781  agcccgagca tatccctgac cctgatgcta agaagcccga ggactgggat gaagagatgg
841  acggagagtg ggaaccccca gtgattcaga accctgagta caagggtgag tggaagcccc
901  ggcagatcga caacccagat tacaagggca cttggatcca cccagaaatt gacaaccccg
961  agtattctcc cgatcccagt atctatgcct atgataactt tggcgtgctg ggcctggacc
1021 tctggcaggt caagtctggc accatctttg acaacttcct catcaccaac gatgaggcat
1081 acgctgagga gtttggcaac gagacgtggg gcgtaacaaa ggcagcagag aaacaaatga
1141 aggacaaaca ggacgaggag cagaggctta aggaggagga agaagacaag aaacgcaaag
1201 aggaggagga ggcagaggac aaggaggatg atgaggacaa agatgaggat gaggaggatg
1261 aggaggacaa ggaggaagat gaggaggaag atgtccccgg ccaggccaag gacgagctgt
1321 agagaggcct gcctccaggg ctggactgag gcctgagcgc tcctgccgca gagcttgccg
1381 cgccaaataa tgtctctgtg agactcgaga actttcattt ttttccaggc tggttcggat
1441 ttggggtgga ttttggtttt gttcccctcc tccactctcc cccacccct ccccgccctt
1501 tttttttttt tttttaaact ggtattttat cctttgattc tccttcagcc ctcacccctg
1561 gttctcatct ttcttgatca acatcttttc ttgcctctgt gccccttctc tcatctctta
1621 gctcccctcc aacctggggg gcagtggtgt ggagaagcca caggcctgag atttcatctg
1681 ctctccttcc tggagcccag aggagggcag cagaaggggg tggtgtctcc aacccccag
1741 cactgaggaa gaacggggct cttctcattt cacccctccc tttctcccct gcccccagga
1801 ctgggccact tctgggtggg gcagtgggtc ccagattggc tcacactgag aatgtaagaa
1861 ctacaaacaa aatttctatt aaattaaatt ttgtgtctc                   1899
```

The amino acid sequence of human CRT protein (SEQ ID NO:2) is shown below

```
1    MLLSVPLLLG LLGLAVAEPA VYFKEQFLDG DGWTSRWIES KHKSDFGKFV LSSGKFYGDE
61   EKDKGLQTSQ DARFYALSAS FEPFSNKGQT LVVQFTVKHE QNIDCGGGYV KLFPNSLDQT
121  DMHGDSEYNI MFGPDICGPG TKKVHVIFNY KGKNVLINKD IRCKDDEFTH LYTLIVRPDN
181  TYEVKIDNSQ VESGSLEDDW DFLPPKKIKD PDASKPEDWD ERAKIDDPTD SKPEDWDKPE
241  HIPDPDAKKP EDWDEEMDGE WEPPVIQNPE YKGEWKPRQI DNPDYKGTWI HPEIDNPEYS
301  PDPSIYAYDN FGVLGLDLWQ VKSGTIFDNF LITNDEAYAE EFGNETWGVT KAAEKQMKDK
361  QDEEQRLKEE EEDKKRKEEE EAEDKEDDED KDEDEEDEED KEEDEEEDVP GQAKDEL     417
```

The structure of polypeptides, peptides, other functional derivatives, including mimetics of CRT are preferably based on structure and amino acid sequence of CRT, preferably human CRT, SEQ ID NO:2 above. (See also, McCauliffe (1990) J. Clin. Invest. 86:332-335; Burns (1994) Nature 367:476-480; Coppolino (1998) Int. J. Biochem. Cell Biol. 30:553-558)

SARS-CoV Genomic Sequences, and Sequences of Polypeptides

The genomic nucleotide sequence of the SARS coronavirus (nt 1 to 29751; SEQ ID NO:3), Tor2 strain , is deposited in Genbank under access no. NC_004718 (available at WWW URL ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=30271926 . See, He, R. *et al.*, *Biochem. Biophys. Res. Commun. 316I* :476-483 (2004) ; Snijder, E.J. *et al.*, *J. Mol. Biol. 331* 991-1004 (2003) ; Marra, MA *et al.*, *Science 300* :1399-1404 (2003). The reference sequence was derived from AY274119. On May 1, 2003 this sequence version replaced gi:30124072.

SEQ ID NO:3

1 atattaggtt tttacctacc caggaaaagc caaccaacct cgatctcttg tagatctgtt
61 ctctaaacga actttaaaat ctgtgtagct gtcgctcggc tgcatgccta gtgcacctac
121 gcagtataaa caataataaa ttttactgtc gttgacaaga aacgagtaac tcgtccctct
181 tctgcagact gcttacggtt tcgtccgtgt tgcagtcgat catcagcata cctaggtttc
241 gtccgggtgt gaccgaaagg taagatggag agccttgttc ttggtgtcaa cgagaaaaca
301 cacgtccaac tcagtttgcc tgtccttcag gttagagacg tgctagtgcg tggcttcggg
361 gactctgtgg aagaggccct atcggaggca cgtgaacacc tcaaaaatgg cacttgtggt
421 ctagtagagc tggaaaaagg cgtactgccc cagcttgaac agccctatgt gttcattaaa
481 cgttctgatg ccttaagcac caatcacggc cacaaggtcg ttgagctggt tgcagaaatg
541 gacggcattc agtacggtcg tagcggtata acactgggag tactcgtgcc acatgtgggc
601 gaaaccccaa ttgcataccg caatgttctt cttcgtaaga acggtaataa gggagccggt
661 ggtcatagct atggcatcga tctaaagtct tatgacttag gtgacgagct tggcactgat
721 cccattgaag attatgaaca aaactggaac actaagcatg gcagtggtgc actccgtgaa
781 ctcactcgtg agctcaatgg aggtgcagtc actcgctatg tcgacaacaa tttctgtggc
841 ccagatgggt accctcttga ttgcatcaaa gattttctcg cacgcgcggg caagtcaatg
901 tgcactcttt ccgaacaact tgattacatc gagtcgaaga gaggtgtcta ctgctgccgt
961 gaccatgagc atgaaattgc ctggttcact gagcgctctg ataagagcta cgagcaccag
1021 acacccttcg aaattaagag tgccaagaaa tttgacactt tcaaagggga atgcccaaag
1081 tttgtgtttc ctcttaactc aaaagtcaaa gtcattcaac cacgtgttga aaagaaaaag
1141 actgagggtt tcatggggcg tatacgctct gtgtaccctg ttgcatctcc acaggagtgt
1201 aacaatatgc acttgtctac cttgatgaaa tgtaatcatt gcgatgaagt ttcatggcag
1261 acgtgcgact ttctgaaagc cacttgtgaa cattgtggca ctgaaaattt agttattgaa
1321 ggacctacta catgtgggta cctacctact aatgctgtag tgaaaatgcc atgtcctgcc
1381 tgtcaagacc cagagattgg acctgagcat agtgttgcag attatcacaa ccactcaaac
1441 attgaaactc gactccgcaa gggaggtagg actagatgtt ttggaggctg tgtgtttgcc
1501 tatgttggct gctataataa gcgtgcctac tgggttcctc gtgctagtgc tgatattggc
1561 tcaggccata ctggcattac tggtgacaat gtggagacct tgaatgagga tctccttgag
1621 atactgagtc gtgaacgtgt taacattaac attgttggcg attttcattt gaatgaagag
1681 gttgccatca ttttggcatc tttctctgct tctacaagtg cctttattga cactataaag
1741 agtcttgatt acaagtcttt caaaaccatt gttgagtcct gcggtaacta taaagttacc
1801 aagggaaagc ccgtaaaagg tgcttggaac attggacaac agagatcagt tttaacacca
1861 ctgtgtggtt ttccctcaca ggctgctggt gttatcagat caatttttgc gcgcacactt
1921 gatgcagcaa accactcaat tcctgatttg caaagagcag ctgtcaccat acttgatggt
1981 atttctgaac agtcattacg tcttgtcgac gccatggttt atacttcaga cctgctcacc
2041 aacagtgtca ttattatggc atatgtaact ggtggtcttg tacaacagac ttctcagtgg
2101 ttgtctaatc ttttgggcac tactgttgaa aaactcaggc ctatctttga atggattgag
2161 gcgaaactta gtgcaggagt tgaatttctc aaggatgctt gggagattct caaatttctc
2221 attacaggtg tttttgacat cgtcaagggt caaatacagg ttgcttcaga taacatcaag
2281 gattgtgtaa aatgcttcat tgatgttgtt aacaaggcac tcgaaatgtg cattgatcaa
2341 gtcactatcg ctggcgcaaa gttgcgatca ctcaacttag gtgaagtctt catcgctcaa
2401 agcaagggac tttaccgtca gtgtatacgt ggcaaggagc agctgcaact actcatgcct
2461 cttaaggcac caaaagaagt aacctttctt gaaggtgatt cacatgacac agtacttacc
2521 tctgaggagg ttgttctcaa gaacggtgaa ctcgaagcac tcgagacgcc cgttgatagc
2581 ttcacaaatg gagctatcgt tggcacacca gtctgtgtaa atggcctcat gctcttagag
2641 attaaggaca aagaacaata ctgcgcattg tctcctggtt tactggctac aaacaatgtc
2701 tttcgcttaa aagggggtgc accaattaaa ggtgtaacct ttggagaaga tactgtttgg
2761 gaagttcaag gttacaagaa tgtgagaatc acatttgagc ttgatgaacg tgttgacaaa
2821 gtgcttaatg aaaagtgctc tgtctacact gttgaatccg gtaccgaagt tactgagttt
2881 gcatgtgttg tagcagaggc tgttgtgaag actttacaac cagtttctga tctccttacc
2941 aacatgggta ttgatcttga tgagtggagt gtagctacat tctacttatt tgatgatgct
3001 ggtgaagaaa acttttcatc acgtatgtat tgttcctttt accctccaga tgaggaagaa
3061 gaggacgatg cagagtgtga ggaagaagaa attgatgaaa cctgtgaaca tgagtacggt

3121 acagaggatg attatcaagg tctccctctg gaatttggtg cctcagctga aacagttcga
3181 gttgaggaag aagaagagga agactggctg gatgatacta ctgagcaatc agagattgag
3241 ccagaaccag aacctacacc tgaagaacca gttaatcagt ttactggtta tttaaaactt
3301 actgacaatg ttgccattaa atgtgttgac atcgttaagg aggcacaaag tgctaatcct
3361 atggtgattg taaatgctgc taacatacac ctgaaacatg gtggtggtgt agcaggtgca
3421 ctcaacaagg caaccaatgg tgccatgcaa aaggagagtg atgattacat taagctaaat
3481 ggccctctta cagtaggagg gtcttgtttg ctttctggac ataatcttgc taagaagtgt
3541 ctgcatgttg ttggacctaa cctaaatgca ggtgaggaca tccagcttct taaggcagca
3601 tatgaaaatt tcaattcaca ggacatctta cttgcaccat tgttgtcagc aggcatattt
3661 ggtgctaaac cacttcagtc tttacaagtg tgcgtgcaga cggttcgtac acaggtttat
3721 attgcagtca atgacaaagc tctttatgag caggttgtca tggattatct tgataacctg
3781 aagcctagag tggaagcacc taaacaagag gagccaccaa acacagaaga ttccaaaact
3841 gaggagaaat ctgtcgtaca gaagcctgtc gatgtgaagc caaaaattaa ggcctgcatt
3901 gatgaggtta ccacaacact ggaagaaact aagtttctta ccaataagtt actcttgttt
3961 gctgatatca atggtaagct ttaccatgat tctcagaaca tgcttagagg tgaagatatg
4021 tctttccttg agaaggatgc accttacatg gtaggtgatg ttatcactag tggtgatatc
4081 acttgtgttg taataccctc caaaaaggct ggtggcacta ctgagatgct ctcaagagct
4141 ttgaagaaag tgccagttga tgagtatata accacgtacc ctggacaagg atgtgctggt
4201 tatacacttg aggaagctaa gactgctctt aagaaatgca aatctgcatt ttatgtacta
4261 ccttcagaag cacctaatgc taaggaagag attctaggaa ctgtatcctg gaatttgaga
4321 gaaatgcttg ctcatgctga agagacaaga aaattaatgc ctatatgcat ggatgttaga
4381 gccataatgg caaccatcca acgtaagtat aaaggaatta aaattcaaga gggcatcgtt
4441 gactatggtg tccgattctt cttttatact agtaaagagc ctgtagcttc tattattacg
4501 aagctgaact ctctaaatga gccgcttgtc acaatgccaa ttggttatgt gacacatggt
4561 tttaatcttg aagaggctgc gcgctgtatg cgttctctta aagctcctgc cgtagtgtca
4621 gtatcatcac cagatgctgt tactacatat aatggatacc tcacttcgtc atcaaagaca
4681 tctgaggagc actttgtaga aacagtttct ttggctggct cttacagaga ttggtcctat
4741 tcaggacagc gtacagagtt aggtgttgaa tttcttaagc gtggtgacaa aattgtgtac
4801 cacactctgg agagccccgt cgagtttcat cttgacggtg aggttctttc acttgacaaa
4861 ctaaagagtc tcttatccct gcgggaggtt aagactataa aagtgttcac aactgtggac
4921 aacactaatc tccacacaca gcttgtggat atgtctatga catatggaca gcagtttggt
4981 ccaacatact tggatggtgc tgatgttaca aaaattaaac ctcatgtaaa tcatgagggt
5041 aagactttct ttgtactacc tagtgatgac acactacgta gtgaagcttt cgagtactac
5101 catactcttg atgagagttt tcttggtagg tacatgtctg ctttaaacca cacaaagaaa
5161 tggaaatttc ctcaagttgg tggtttaact tcaattaaat gggctgataa caattgttat
5221 ttgtctagtg ttttattagc acttcaacag cttgaagtca aattcaatgc accagcactt
5281 caagaggctt attatagagc ccgtgctggt gatgctgcta acttttgtgc actcatactc
5341 gcttacagta ataaaactgt tggcgagctt ggtgatgtca gagaaactat gacccatctt
5401 ctacagcatg ctaatttgga atctgcaaag cgagttctta atgtggtgtg taaacattgt
5461 ggtcagaaaa ctactacctt aacgggtgta gaagctgtga tgtatatggg tactctatct
5521 tatgataatc ttaagacagg tgtttccatt ccatgtgtgt gtggtcgtga tgctacacaa
5581 tatctagtac aacaagagtc ttcttttgtt atgatgtctg caccacctgc tgagtataaa
5641 ttacagcaag gtacattctt atgtgcgaat gagtacactg gtaactatca gtgtggtcat
5701 tacactcata taactgctaa ggagaccctc tatcgtattg acggagctca ccttacaaag
5761 atgtcagagt acaaaggacc agtgactgat gttttctaca aggaaacatc ttacactaca
5821 accatcaagc ctgtgtcgta taaactcgat ggagttactt acacagagat tgaaccaaaa
5881 ttggatgggt attataaaaa ggataatgct tactatacag agcagcctat agaccttgta
5941 ccaactcaac cattaccaaa tgcgagtttt gataatttca aactcacatg ttctaacaca
6001 aaatttgctg atgatttaaa tcaaatgaca ggcttcacaa agccagcttc acgagagcta
6061 tctgtcacat tcttcccaga cttgaatggc gatgtagtgg ctattgacta tagacactat
6121 tcagcgagtt tcaagaaagg tgctaaatta ctgcataagc caattgtttg gcacattaac
6181 caggctacaa ccaagacaac gttcaaacca aacacttggt gtttacgttg tctttggagt
6241 acaaagccag tagatacttc aaattcattt gaagttctgg cagtagaaga cacacaagga
6301 atggacaatc ttgcttgtga aagtcaacaa cccacctctg aagaagtagt ggaaaatcct
6361 accatacaga aggaagtcat agagtgtgac gtgaaaacta ccgaagttgt aggcaatgtc
6421 atacttaaac catcagatga aggtgttaaa gtaacacaag agttaggtca tgaggatctt
6481 atggctgctt atgtggaaaa cacaagcatt accattaaga aacctaatga gctttcacta
6541 gccttaggtt taaaaacaat tgccactcat ggtattgctg caattaatag tgttccttgg
6601 agtaaaattt tggcttatgt caaaccattc ttaggacaag cagcaattac aacatcaaat
6661 tgcgctaaga gattagcaca acgtgtgttt aacaattata tgccttatgt gtttacatta
6721 ttgttccaat tgtgtacttt tactaaaagt accaattcta gaattagagc ttcactacct
6781 acaactattg ctaaaaatag tgttaagagt gttgctaaat tatgtttgga tgccggcatt
6841 aattatgtga agtcacccaa attttctaaa ttgttcacaa tcgctatgtg gctattgttg
6901 ttaagtattt gcttaggttc tctaatctgt gtaactgctg cttttggtgt actcttatct
6961 aattttggtg ctccttctta ttgtaatggc gttagagaat tgtatcttaa ttcgtctaac
7021 gttactacta tggatttctg tgaaggttct tttccttgca gcatttgttt aagtggatta

```
7081   gactcccttg attcttatcc agctcttgaa accattcagg tgacgatttc atcgtacaag
7141   ctagacttga caattttagg tctggccgct gagtgggttt tggcatatat gttgttcaca
7201   aaattctttt atttattagg tctttcagct ataatgcagg tgttctttgg ctattttgct
7261   agtcatttca tcagcaattc ttggctcatg tggtttatca ttagtattgt acaaatggca
7321   cccgtttctg caatggttag gatgtacatc ttctttgctt ctttctacta catatggaag
7381   agctatgttc atatcatgga tggttgcacc tcttcgactt gcatgatgtg ctataagcgc
7441   aatcgtgcca cacgcgttga gtgtacaact attgttaatg gcatgaagag atctttctat
7501   gtctatgcaa atggaggccg tggcttctgc aagactcaca attggaattg tctcaattgt
7561   gacacatttt gcactggtag tacattcatt agtgatgaag ttgctcgtga tttgtcactc
7621   cagtttaaaa gaccaatcaa ccctactgac cagtcatcgt atattgttga tagtgttgct
7681   gtgaaaaatg gcgcgcttca cctctacttt gacaaggctg gtcaaaagac ctatgagaga
7741   catccgctct cccattttgt caatttagac aatttgagag ctaacaacac taaaggttca
7801   ctgcctatta atgtcatagt tttgatggc aagtccaaat gcgacgagtc tgcttctaag
7861   tctgcttctg tgtactacag tcagctgatg tgccaaccta ttctgttgct tgaccaagct
7921   cttgtatcag acgttggaga tagtactgaa gtttccgtta agatgtttga tgcttatgtc
7981   gacacctttt cagcaacttt tagtgttcct atggaaaaac ttaaggcact tgttgctaca
8041   gctcacagcg agttagcaaa gggtgtagct ttagatggtg tcctttctac attcgtgtca
8101   gctgcccgac aaggtgttgt tgataccgat gttgacacaa aggatgttat tgaatgtctc
8161   aaactttcac atcactctga cttagaagtg acaggtgaca gttgtaacaa tttcatgctc
8221   acctataata aggttgaaaa catgacgccc agagatcttg gcgcatgtat tgactgtaat
8281   gcaaggcata tcaatgccca agtagcaaaa agtcacaatg tttcactcat ctggaatgta
8341   aaagactaca tgtctttatc tgaacagctg cgtaaacaaa ttcgtagtgc tgccaagaag
8401   aacaacatac cttttagact aacttgtgct acaactagac aggttgtcaa tgtcataact
8461   actaaaatct cactcaaggg tggtaagatt gttagtactt gttttaaact tatgcttaag
8521   gccacattat tgtgcgttct tgctgcattg gtttgttata tcgttatgcc agtacataca
8581   ttgtcaatcc atgatggtta cacaaatgaa atcattggtt acaaagccat tcaggatggt
8641   gtcactcgtg acatcatttc tactgatgat tgttttgcaa ataaacatgc tggttttgac
8701   gcatggttta gccagcgtgg tggttcatac aaaaatgaca aaagctgccc tgtagtagct
8761   gctatcatta caagagagat tggtttcata gtgcctggct taccgggtac tgtgctgaga
8821   gcaatcaatg gtgacttctt gcattttcta cctcgtgttt ttagtgctgt tggcaacatt
8881   tgctacacac cttccaaact cattgagtat agtgattttg ctacctctgc ttgcgttctt
8941   gctgctgagt gtacaatttt taaggatgct atgggcaaac ctgtgccata ttgttatgac
9001   actaatttgc tagagggttc tatttcttat agtgagcttc gtccagacac tcgttatgtg
9061   cttatggatg gttccatcat acagtttcct aacacttacc tggagggttc tgttagagta
9121   gtaacaactt ttgatgctga gtactgtaga catggtacat gcgaaaggtc agaagtaggt
9181   atttgcctat ctaccagtgg tagatgggtt cttaataatg agcattacag agctctatca
9241   ggagttttct gtggtgttga tgcgatgaat ctcatagcta acatctttac tcctcttgtg
9301   caacctgtgg gtgctttaga tgtgtctgct tcagtagtgg ctggtggtat tattgccata
9361   ttggtgactt gtgctgccta ctactttatg aaattcagac gtgtttttgg tgagtacaac
9421   catgttgttg ctgctaatgc acttttgttt ttgatgtctt tcactatact ctgtctggta
9481   ccagcttaca gctttctgcc gggagtctac tcagtctttt acttgtactt gacattctat
9541   ttcaccaatg atgtttcatt cttggctcac cttcaatggt ttgccatgtt ttctcctatt
9601   gtgccttttt ggataacagc aatctatgta ttctgtattt ctctgaagca ctgccattgg
9661   ttctttaaca actatcttag gaaaagagtc atgtttaatg gagttacatt tagtaccttc
9721   gaggaggctg ctttgtgtac ctttttgctc aacaaggaaa tgtacctaaa attgcgtagc
9781   gagacactgt tgccacttac acagtataac aggtatcttg ctctatataa caagtacaag
9841   tatttcagtg gagccttaga tactaccagc tatcgtgaag cagcttgctg ccacttagca
9901   aaggctctaa atgactttag caactcaggt gctgatgttc tctaccaacc accacagaca
9961   tcaatcactt ctgctgttct gcagagtggt tttaggaaaa tggcattccc gtcaggcaaa
10021  gttgaagggt gcatggtaca agtaacctgt ggaactacaa ctcttaatgg attgtggttg
10081  gatgacacag tatactgtcc aagacatgtc atttgcacag cagaagacat gcttaatcct
10141  aactatgaag atctgctcat tcgcaaatcc aaccatagct ttcttgttca ggctggcaat
10201  gttcaacttc gtgttattgg ccattctatg caaaattgtc tgcttaggct taaagttgat
10261  acttctaacc ctaagacacc caagtataaa tttgtccgta tccaacctgg tcaaacattt
10321  tcagttctag catgctacaa tggttcacca tctggtgttt atcagtgtgc catgagacct
10381  aatcatacca ttaaaggttc tttccttaat ggatcatgtg gtagtgttgg ttttaacatt
10441  gattatgatt gcgtgtcttt ctgctatatg catcatatgg agcttccaac aggagtacac
10501  gctggtactg acttagaagg taaattctat ggtccatttg ttgacagaca aactgcacag
10561  gctgcaggta cagacacaac cataacatta aatgttttgg catggctgta tgctgctgtt
10621  atcaatggtg ataggtggtt tcttaataga ttcaccacta ctttgaatga ctttaacctt
10681  gtggcaatga agtacaacta tgaacctttg acacaagatc atgttgacat attgggacct
10741  ctttctgctc aaacaggaat tgccgtctta gatatgtgtg ctgctttgaa agagctgctg
10801  cagaatggta tgaatggtcg tactatcctt ggtagcacta ttttagaaga tgagtttaca
10861  ccatttgatg ttgttagaca atgctctggt gttaccttcc aaggtaagtt caagaaaatt
10921  gttaagggca ctcatcattg gatgctttta actttcttga catcactatt gattcttgtt
10981  caaagtacac agtggtcact gtttttcttt gtttacgaga atgctttctt gccatttact
```

11041	cttggtatta	tggcaattgc	tgcatgtgct	atgctgcttg	ttaagcataa	gcacgcattc
11101	ttgtgcttgt	ttctgttacc	ttctcttgca	acagttgctt	actttaatat	ggtctacatg
11161	cctgctagct	gggtgatgcg	tatcatgaca	tggcttgaat	tggctgacac	tagcttgtct
11221	ggttataggc	ttaaggattg	tgttatgtat	gcttcagctt	tagttttgct	tattctcatg
11281	acagctcgca	ctgtttatga	tgatgctgct	agacgtgttt	ggacactgat	gaatgtcatt
11341	acacttgttt	acaaagtcta	ctatggtaat	gctttagatc	aagctatttc	catgtgggcc
11401	ttagttattt	ctgtaacctc	taactattct	ggtgtcgtta	cgactatcat	gtttttagct
11461	agagctatag	tgtttgtgtg	tgttgagtat	tacccattgt	tatttattac	tggcaacacc
11521	ttacagtgta	tcatgcttgt	ttattgtttc	ttaggctatt	gttgctgctg	ctactttggc
11581	cttttctgtt	tactcaaccg	ttacttcagg	cttactcttg	gtgtttatga	ctacttggtc
11641	tctacacaag	aatttaggta	tatgaactcc	caggggcttt	tgcctcctaa	gagtagtatt
11701	gatgctttca	agcttaacat	taagttgttg	ggtattggag	gtaaaccatg	tatcaaggtt
11761	gctactgtac	agtctaaaat	gtctgacgta	aagtgcacat	ctgtggtact	gctctcggtt
11821	cttcaacaac	ttagagtaga	gtcatcttct	aaattgtggg	cacaatgtgt	acaactccac
11881	aatgatattc	ttcttgcaaa	agacacaact	gaagctttcg	agaagatggt	ttctcttttg
11941	tctgttttgc	tatccatgca	gggtgctgta	gacattaata	ggttgtgcga	ggaaatgctc
12001	gataaccgtg	ctactcttca	ggctattgct	tcagaattta	gttctttacc	atcatatgcc
12061	gcttatgcca	ctgcccagga	ggcctatgag	caggctgtag	ctaatggtga	ttctgaagtc
12121	gttctcaaaa	agttaaagaa	atctttgaat	gtggctaaat	ctgagtttga	ccgtgatgct
12181	gccatgcaac	gcaagttgga	aaagatggca	gatcaggcta	tgacccaaat	gtacaaacag
12241	gcaagatctg	aggacaagag	ggcaaaagta	actagtgcta	tgcaaacaat	gctcttcact
12301	atgcttagga	agcttgataa	tgatgcactt	aacaacatta	tcaacaatgc	gcgtgatggt
12361	tgtgttccac	tcaacatcat	accattgact	acagcagcca	aactcatggt	tgttgtccct
12421	gattatggta	cctacaagaa	cacttgtgat	ggtaacacct	ttacatatgc	atctgcactc
12481	tgggaaatcc	agcaagttgt	tgatgcggat	agcaagattg	ttcaacttag	tgaaattaac
12541	atggacaatt	caccaaattt	ggcttggcct	cttattgtta	cagctctaag	agccaactca
12601	gctgttaaac	tacagaataa	tgaactgagt	ccagtagcac	tacgacagat	gtcctgtgcg
12661	gctggtacca	cacaaacagc	ttgtactgat	gacaatgcac	ttgcctacta	taacaattcg
12721	aagggaggta	ggtttgtgct	ggcattacta	tcagaccacc	aagatctcaa	atgggctaga
12781	ttccctaaga	gtgatggtac	aggtacaatt	tacacagaac	tggaaccacc	ttgtaggttt
12841	gttacagaca	caccaaaagg	gcctaaagtg	aaatacttgt	acttcatcaa	aggcttaaac
12901	aacctaaata	gaggtatggt	gctgggcagt	ttagctgcta	cagtacgtct	tcaggctgga
12961	aatgctacag	aagtacctgc	caattcaact	gtgctttcct	tctgtgcttt	tgcagtagac
13021	cctgctaaag	catataagga	ttacctagca	agtggaggac	aaccaatcac	caactgtgtg
13081	aagatgttgt	gtacacacac	tggtacagga	caggcaatta	ctgtaacacc	agaagctaac
13141	atggaccaag	agtcctttgg	tggtgcttca	tgttgtctgt	attgtagatg	ccacattgac
13201	catccaaatc	ctaaaggatt	ctgtgacttg	aaaggtaagt	acgtccaaat	acctaccact
13261	tgtgctaatg	acccagtggg	ttttacactt	agaaacacag	tctgtaccgt	ctgcggaatg
13321	tggaaaggtt	atggctgtag	ttgtgaccaa	ctccgcgaac	ccttgatgca	gtctgcggat
13381	gcatcaacgt	tttaaacgg	gtttgcggtg	taagtgcagc	ccgtcttaca	ccgtgcggca
13441	caggcactag	tactgatgtc	gtctacaggg	cttttgatat	ttacaacgaa	aaagttgctg
13501	gttttgcaaa	gttcctaaaa	actaattgct	gtcgcttcca	ggagaaggat	gaggaaggca
13561	atttattaga	ctcttacttt	gtagttaaga	ggcatactat	gtctaactac	caacatgaag
13621	agactattta	taacttggtt	aaagattgtc	cagcggttgc	tgtccatgac	tttttcaagt
13681	ttagagtaga	tggtgacatg	gtaccacata	tatcacgtca	gcgtctaact	aaatacacaa
13741	tggctgattt	agtctatgct	ctacgtcatt	ttgatgaggg	taattgtgat	acattaaaag
13801	aaatactcgt	cacatacaat	tgctgtgatg	atgattattt	caataagaag	gattggtatg
13861	acttcgtaga	gaatcctgac	atcttacgcg	tatatgctaa	cttaggtgag	cgtgtacgcc
13921	aatcattatt	aaagactgta	caattctgcg	atgctatgcg	tgatgcaggc	attgtaggcg
13981	tactgacatt	agataatcag	gatcttaatg	ggaactggta	cgatttcggt	gatttcgtac
14041	aagtagcacc	aggctgcgga	gttcctattg	tggattcata	ttactcattg	ctgatgccca
14101	tcctcacttt	gactagggca	ttggctgctg	agtcccatat	ggatgctgat	ctcgcaaaac
14161	cacttattaa	gtgggatttg	ctgaaatatg	attttacgga	agagagactt	tgtctcttcg
14221	accgttattt	taaatattgg	gaccagacat	accatcccaa	ttgtattaac	tgtttggatg
14281	ataggtgtat	ccttcattgt	gcaaacttta	atgtgttatt	ttctactgtg	tttccaccta
14341	caagttttgg	accactagta	agaaaaatat	ttgtagatgg	tgttcctttt	gttgtttcaa
14401	ctggatacca	ttttcgtgag	ttaggagtcg	tacataatca	ggatgtaaac	ttacatagct
14461	cgcgtctcag	tttcaaggaa	cttttagtgt	atgctgctga	tccagctatg	catgcagctt
14521	ctggcaattt	attgctagat	aaacgcacta	catgcttttc	agtagctgca	ctaacaaaca
14581	atgttgcttt	tcaaactgtc	aaacccggta	attttaataa	agacttttat	gactttgctg
14641	tgtctaaagg	tttctttaag	gaaggaagtt	ctgttgaact	aaaacacttc	ttctttgctc
14701	aggatggcaa	cgctgctatc	agtgattatg	actattatcg	ttataatctg	ccaacaatgt
14761	gtgatatcag	acaactccta	ttcgtagttg	aagttgttga	taaatacttt	gattgttacg
14821	atggtggctg	tattaatgcc	aaccaagtaa	tcgttaacaa	tctggataaa	tcagctggtt
14881	tcccatttaa	taaatggggt	aaggctagac	tttattatga	ctcaatgagt	tatgaggatc
14941	aagatgcact	tttcgcgtat	actaagcgta	atgtcatccc	tactataact	caaatgaatc

15001 ttaagtatgc cattagtgca aagaatagag ctcgcaccgt agctggtgtc tctatctgta
15061 gtactatgac aaatagacag tttcatcaga aattattgaa gtcaatagcc gccactagag
15121 gagctactgt ggtaattgga acaagcaagt tttacggtgg ctggcataat atgttaaaaa
15181 ctgtttacag tgatgtagaa actccacacc ttatgggttg ggattatcca aaatgtgaca
15241 gagccatgcc taacatgctt aggataatgg cctctcttgt tcttgctcgc aaacataaca
15301 cttgctgtaa cttatcacac cgtttctaca ggttagctaa cgagtgtgcg caagtattaa
15361 gtgagatggt catgtgtggc ggctcactat atgttaaacc aggtggaaca tcatccggtg
15421 atgctacaac tgcttatgct aatagtgtct ttaacatttg tcaagctgtt acagccaatg
15481 taaatgcact tctttcaact gatggtaata agatagctga caagtatgtc cgcaatctac
15541 aacacaggct ctatgagtgt ctctatagaa atagggatgt tgatcatgaa ttcgtggatg
15601 agttttacgc ttacctgcgt aaacatttct ccatgatgat tctttctgat gatgccgttg
15661 tgtgctataa cagtaactat gcggctcaag gtttagtagc tagcattaag aactttaagg
15721 cagttcttta ttatcaaaat aatgtgttca tgtctgaggc aaaatgttgg actgagactg
15781 accttactaa aggacctcac gaattttgct cacagcatac aatgctagtt aaacaaggag
15841 atgattacgt gtacctgcct tacccagatc catcaagaat attaggcgca ggctgttttg
15901 tcgatgatat tgtcaaaaca gatggtacac ttatgattga aaggttcgtg tcactggcta
15961 ttgatgctta cccacttaca aaacatccta atcaggagta tgctgatgtc tttcacttgt
16021 atttacaata cattagaaag ttacatgatg agcttactgg ccacatgttg gacatgtatt
16081 ccgtaatgct aactaatgat aacacctcac ggtactggga acctgagttt tatgaggcta
16141 tgtacacacc acatacagtc ttgcaggctg taggtgcttg tgtattgtgc aattcacaga
16201 cttcacttcg ttgcggtgcc tgtattagga gaccattcct atgttgcaag tgctgctatg
16261 accatgtcat ttcaacatca cacaaattag tgttgtctgt taatccctat gtttgcaatg
16321 ccccaggttg tgatgtcact gatgtgacac aactgtatct aggaggtatg agctattatt
16381 gcaagtcaca taagcctccc attagttttc cattatgtgc taatggtcag gtttttggtt
16441 tatacaaaaa cacatgtgta ggcagtgaca atgtcactga cttcaatgcg atagcaacat
16501 gtgattggac taatgctggc gattacatac ttgccaacac ttgtactgag agactcaagc
16561 ttttcgcagc agaaacgctc aaagccactg aggaaacatt taagctgtca tatggtattg
16621 ccactgtacg cgaagtactc tctgacagag aattgcatct ttcatgggag gttggaaaac
16681 ctagaccacc attgaacaga aactatgtct ttactggtta ccgtgtaact aaaaatagta
16741 aagtacagat tggagagtac acctttgaaa aaggtgacta tggtgatgct gttgtgtaca
16801 gaggtactac gacatacaag ttgaatgttg gtgattactt tgtgttgaca tctcacactg
16861 taatgccact tagtgcacct actctagtgc cacaagagca ctatgtgaga attactggct
16921 tgtacccaac actcaacatc tcagatgagt tttctagcaa tgttgcaaat tatcaaaagg
16981 tcggcatgca aaagtactct acactccaag gaccacctgg tactggtaag agtcattttg
17041 ccatcggact tgctctctat tacccatctg ctcgcatagt gtatacggca tgctctcatg
17101 cagctgttga tgccctatgt gaaaaggcat taaaatattt gcccatagat aaatgtagta
17161 gaatcatacc tgcgcgtgcg cgcgtagagt gttttgataa attcaaagtg aattcaacac
17221 tagaacagta tgttttctgc actgtaaatg cattgccaga aacaactgct gacattgtag
17281 tctttgatga aatctctatg gctactaatt atgacttgag tgttgtcaat gctagacttc
17341 gtgcaaaaca ctacgtctat attggcgatc ctgctcaatt accagccccc cgcacattgc
17401 tgactaaagg cacactagaa ccagaatatt ttaattcagt gtgcagactt atgaaaacaa
17461 taggtccaga catgttcctt ggaacttgtc gccgttgtcc tgctgaaatt gttgacactg
17521 tgagtgcttt agtttatgac aataagctaa aagcacacaa ggataagtca gctcaatgct
17581 tcaaaatgtt ctacaaaggt gttattacac atgatgtttc atctgcaatc aacagacctc
17641 aaataggcgt tgtaagagaa tttcttacac gcaatcctgc ttggagaaaa gctgttttta
17701 tctcacctta taattcacag aacgctgtag cttcaaaaat cttaggattg cctacgcaga
17761 ctgttgattc atcacagggt tctgaatatg actatgtcat attcacacaa actactgaaa
17821 cagcacactc ttgtaatgtc aaccgcttca atgtggctat cacaagggca aaaattggca
17881 ttttgtgcat aatgtctgat agagatcttt atgacaaact gcaatttaca agtctagaaa
17941 taccacgtcg caatgtggct acattacaag cagaaaatgt aactggactt tttaaggact
18001 gtagtaagat cattactggt cttcatccta cacaggcacc tacacacctc agcgttgata
18061 taaagttcaa gactgaagga ttatgtgttg acataccagg cataccaaag gacatgacct
18121 accgtagact catctctatg atgggtttca aaatgaatta ccaagtcaat ggttacccta
18181 atatgtttat cacccgcgaa gaagctattc gtcacgttcg tgcgtggatt ggctttgatg
18241 tagagggctg tcatgcaact agagatgctg tgggtactaa cctacctctc cagctaggat
18301 tttctacagg tgttaactta gtagctgtac cgactggtta tgttgacact gaaaataaca
18361 cagaattcac cagagttaat gcaaaacctc caccaggtga ccagtttaaa catcttatac
18421 cactcatgta taaaggcttg ccctggaatg tagtgcgtat taagatagta caaatgctca
18481 gtgatacact gaaaggattg tcagacagag tcgtgttcgt cctttgggcg catggctttg
18541 agcttacatc aatgaagtac tttgtcaaga ttggacctga aagaacgtgt tgtctgtgtg
18601 acaaacgtgc aacttgcttt tctacttcat cagatactta tgcctgctgg aatcattctg
18661 tgggttttga ctatgtctat aacccattta tgattgatgt tcagcagtgg ggctttacgg
18721 gtaaccttca gagtaaccat gaccaacatt gccaggtaca tggaaatgca catgtggcta
18781 gttgtgatgc tatcatgact agatgtttag cagtccatga gtgctttgtt aagcgcgttg
18841 attggtctgt tgaatacccc attataggag atgaactgag ggttaattct gcttgcagaa
18901 aagtacaaca catggttgtg aagtctgcat tgcttgctga taagtttcca gttcttcatg

18961 acattggaaa tccaaaggct atcaagtgtg tgcctcaggc tgaagtagaa tggaagttct
19021 acgatgctca gccatgtagt gacaaagctt acaaaataga ggaactcttc tattcttatg
19081 ctacacatca cgataaattc actgatggtg tttgtttgtt ttggaattgt aacgttgatc
19141 gttacccagc caatgcaatt gtgtgtaggt ttgacacaag agtcttgtca aacttgaact
19201 taccaggctg tgatggtggt agtttgtatg tgaataagca tgcattccac actccagctt
19261 tcgataaaag tgcatttact aatttaaagc aattgccttt cttttactat tctgatagtc
19321 cttgtgagtc tcatggcaaa caagtagtgt cggatattga ttatgttcca ctcaaatctg
19381 ctacgtgtat tacacgatgc aatttaggtg gtgctgtttg cagacaccat gcaaatgagt
19441 accgacagta cttggatgca tataatatga tgatttctgc tggatttagc ctatggattt
19501 acaaacaatt tgatacttat aacctgtgga atacatttac caggttacag agtttagaaa
19561 atgtggctta taatgttgtt aataaaggac actttgatgg acacgccggc gaagcacctg
19621 tttccatcat taataatgct gtttacacaa aggtagatgg tattgatgtg gagatctttg
19681 aaaataagac aacacttcct gttaatgttg catttgagct ttgggctaag cgtaacatta
19741 aaccagtgcc agagattaag atactcaata atttgggtgt tgatatcgct gctaatactg
19801 taatctggga ctacaaaaga gaagccccag cacatgtatc tacaataggt gtctgcacaa
19861 tgactgacat tgccaagaaa cctactgaga gtgcttgttc ttcacttact gtcttgtttg
19921 atggtagagt ggaaggacag gtagaccttt ttagaaacgc ccgtaatggt gttttaataa
19981 cagaaggttc agtcaaaggt ctaacacctt caaagggacc agcacaagct agcgtcaatg
20041 gagtcacatt aattggagaa tcagtaaaaa cacagtttaa ctactttaag aaagtagacg
20101 gcattattca acagttgcct gaaacctact ttactcagag cagagactta gaggatttta
20161 agcccagatc acaaatggaa actgactttc tcgagctcgc tatggatgaa ttcatacagc
20221 gatataagct cgagggctat gccttcgaac acatcgttta tggagatttc agtcatggac
20281 aacttggcgg tcttcattta atgataggct tagccaagcg ctcacaagat tcaccactta
20341 aattagagga ttttatccct atggacagca cagtgaaaaa ttacttcata acagatgcgc
20401 aaacaggttc atcaaaatgt gtgtgttctg tgattgatct tttacttgat gactttgtcg
20461 agataataaa gtcacaagat ttgtcagtga tttcaaaagt ggtcaaggtt acaattgact
20521 atgctgaaat ttcattcatg ctttggtgta aggatggaca tgttgaaacc ttctacccaa
20581 aactacaagc aagtcaagcg tggcaaccag gtgttgcgat gcctaacttg tacaagatgc
20641 aaagaatgct tcttgaaaag tgtgaccttc agaattatgg tgaaaatgct gttataccaa
20701 aaggaataat gatgaatgtc gcaaagtata ctcaactgtg tcaatactta aatacactta
20761 ctttagctgt accctacaac atgagagtta ttcactttgg tgctggctct gataaaggag
20821 ttgcaccagg tacagctgtg ctcagacaat ggttgccaac tggcacacta cttgtcgatt
20881 cagatcttaa tgacttcgtc tccgacgcag attctacttt aattggagac tgtgcaacag
20941 tacatacggc taataaatgg gaccttatta ttagcgatat gtatgaccct aggaccaaac
21001 atgtgacaaa agagaatgac tctaaagaag ggtttttcac ttatctgtgt ggatttataa
21061 agcaaaaact agccctgggt ggttctatag ctgtaaagat aacagagcat tcttggaatg
21121 ctgaccttta caagcttatg ggccatttct catggtggac agcttttgtt acaaatgtaa
21181 atgcatcatc atcggaagca tttttaattg ggctaacta tcttggcaag ccgaaggaac
21241 aaattgatgg ctataccatg catgctaact acattttctg gaggaacaca aatcctatcc
21301 agttgtcttc ctattcactc tttgacatga gcaaatttcc tcttaaatta agaggaactg
21361 ctgtaatgtc tcttaaggag aatcaaatca atgatatgat ttattctctt ctggaaaaag
21421 gtaggcttat cattagagaa aacaacagag ttgtggtttc aagtgatatt cttgttaaca

Gene S underscored→

21481 actaaacgaa c**ATG**tttatt ttcttattat ttcttactct cactagtggt agtgaccttg
21541 accggtgcac cacttttgat gatgttcaag ctcctaatta cactcaacat acttcatcta
21601 tgaggggggt ttactatcct gatgaaattt ttagatcaga cactctttat ttaactcagg
21661 atttatttct tccattttat tctaatgtta cagggtttca tactattaat catacgtttg
21721 gcaaccctgt catacctttt aaggatggta tttattttgc tgccacagag aaatcaaatg
21781 ttgtccgtgg ttgggttttt ggttctacca tgaacaacaa gtcacagtcg gtgattatta
21841 ttaacaattc tactaatgtt gttatacgag catgtaactt tgaattgtgt gacaaccctt
21901 tctttgctgt ttctaaaccc atgggtacac agacacatac tatgatattc gataatgcat
21961 ttaattgcac tttcgagtac atatctgatg ccttttcgct tgatgtttca gaaaagtcag
22021 gtaattttaa acacttacga gagtttgtgt ttaaaaataa agatgggttt ctctatgttt
22081 ataagggcta tcaacctata gatgtagttc gtgatctacc ttctggtttt aacactttga
22141 aacctatttt taagttgcct cttggtatta acattacaaa ttttagagcc attcttacag
22201 ccttttcacc tgctcaagac atttggggca cgtcagctgc agcctatttt gttggctatt
22261 taaagccaac tacatttatg ctcaagtatg atgaaaatgg tacaatcaca gatgctgttg
22321 attgttctca aaatccactt gctgaactca aatgctctgt taagagcttt gagattgaca
22381 aaggaattta ccagacctct aatttcaggg ttgttccctc aggagatgtt gtgagattcc
22441 ctaatattac aaacttgtgt ccttttggag aggtttttaa tgctactaaa ttcccttctg
22501 tctatgcatg ggagagaaaa aaaatttcta attgtgttgc tgattactct gtgctctaca
22561 actcaacatt tttttcaacc tttaagtgct atggcgtttc tgccactaag ttgaatgatc
22621 tttgcttctc caatgtctat gcagattctt ttgtagtcaa gggagatgat gtaagacaaa
22681 tagcgccagg acaaactggt gttattgctg attataatta taaattgcca gatgatttca
22741 tgggttgtgt ccttgcttgg aatactagga acattgatgc tacttcaact ggtaattata
22801 attataaata taggtatctt agacatggca agcttaggcc ctttgagaga gacatatcta

22861 atgtgccttt ctcccctgat ggcaaacctt gcacccccacc tgctcttaat tgttattggc
22921 cattaaatga ttatggtttt tacaccacta ctggcattgg ctaccaacct tacagagttg
22981 tagtactttc ttttgaactt ttaaatgcac cggccacggt ttgtggacca aaattatcca
23041 ctgaccttat taagaaccag tgtgtcaatt ttaattttaa tggactcact ggtactggtg
23101 tgttaactcc ttcttcaaag agatttcaac catttcaaca atttggccgt gatgtttctg
23161 atttcactga ttccgttcga gatcctaaaa catctgaaat attagacatt tcaccttgcg
23221 cttttggggg tgtaagtgta attacacctg gaacaaatgc ttcatctgaa gttgctgttc
23281 tatatcaaga tgttaactgc actgatgttt ctacagcaat tcatgcagat caactcacac
23341 cagcttggcg catatattct actggaaaca atgtattcca gactcaagca ggctgtctta
23401 taggagctga gcatgtcgac acttcttatg agtgcgacat tcctattgga gctggcattt
23461 gtgctagtta ccatacagtt tctttattac gtagtactag ccaaaaatct attgtggctt
23521 atactatgtc tttaggtgct gatagttcaa ttgcttactc taataacacc attgctatac
23581 ctactaactt ttcaattagc attactacag aagtaatgcc tgtttctatg gctaaaacct
23641 ccgtagattg taatatgtac atctgcggag attctactga atgtgctaat ttgcttctcc
23701 aatatggtag cttttgcaca caactaaatc gtgcactctc aggtattgct gctgaacagg
23761 atcgcaacac acgtgaagtg ttcgctcaag tcaaacaaat gtacaaaacc ccaactttga
23821 aatattttgg tggttttaat ttttcacaaa tattacctga ccctctaaag ccaactaaga
23881 ggtcttttat tgaggacttg ctctttaata aggtgacact cgctgatgct ggcttcatga
23941 agcaatatgg cgaatgccta ggtgatatta atgctagaga tctcatttgt gcgcagaagt
24001 tcaatggact tacagtgttg ccacctctgc tcactgatga tatgattgct gcctacactg
24061 ctgctctagt tagtggtact gccactgctg gatggacatt tggtgctggc gctgctcttc
24121 aaataccttt tgctatgcaa atggcatata ggttcaatgg cattggagtt acccaaaatg
24181 ttctctatga gaaccaaaaa caaatcgcca accaatttaa caaggcgatt agtcaaattc
24241 aagaatcact tacaacaaca tcaactgcat tgggcaagct gcaagacgtt gttaaccaga
24301 atgctcaagc attaaacaca cttgttaaac aacttagctc taattttggt gcaatttcaa
24361 gtgtgctaaa tgatatcctt tcgcgacttg ataaagtcga ggcggaggta caaattgaca
24421 ggttaattac aggcagactt caaagccttc aaacctatgt aacacaacaa ctaatcaggg
24481 ctgctgaaat caggggcttct gctaatcttg ctgctactaa aatgtctgag tgtgttcttg
24541 gacaatcaaa aagagttgac ttttgtggaa agggctacca ccttatgtcc ttcccacaag
24601 cagccccgca tggtgttgtc ttcctacatg tcacgtatgt gccatcccag gagaggaact
24661 tcaccacagc gccagcaatt tgtcatgaag gcaaagcata cttccctcgt gaaggtgttt
24721 ttgtgtttaa tggcacttct tggtttatta cacagaggaa cttcttttct ccacaaataa
24781 ttactacaga caatacattt gtctcaggaa attgtgatgt cgttattggc atcattaaca
24841 acacagttta tgatcctctg caacctgagc ttgactcatt caaagaagag ctggacaagt
24901 acttcaaaaa tcatacatca ccagatgttg atcttggcga catttcaggc attaacgctt
24961 ctgtcgtcaa cattcaaaaa gaaattgacc gcctcaatga ggtcgctaaa aatttaaatg
25021 aatcactcat tgaccttcaa gaattgggaa aatatgagca atatattaaa tggccttggt
25081 atgtttggct cggcttcatt gctggactaa ttgccatcgt catggttaca atcttgcttt
25141 gttgcatgac tagttgttgc agttgcctca agggtgcatg ctcttgtggt tcttgctgca
25201 agtttgatga ggatgactct gagccagttc tcaagggtgt caaattacat tacaca*TAA*a
25261 cgaacttatg gatttgttta tgagattttt tactcttaga tcaattactg cacagccagt
25321 aaaaattgac aatgcttctc ctgcaagtac tgttcatgct acagcaacga taccgctaca
25381 agcctcactc cctttcggat ggcttgttat tggcgttgca tttcttgctg tttttcagag
25441 cgctaccaaa ataattgcgc tcaataaaag atggcagcta gccctttata agggcttcca
25501 gttcatttgc aatttactgc tgctatttgt taccatctat tcacatcttt tgcttgtcgc
25561 tgcaggtatg gaggcgcaat ttttgtacct ctatgccttg atatatttc tacaatgcat
25621 caacgcatgt agaattatta tgagatgttg gctttgttgg aagtgcaaat ccaagaaccc
25681 attactttat gatgccaact actttgtttg ctggcacaca cataactatg actactgtat
25741 accatataac agtgtcacag atacaattgt cgttactgaa ggtgacggca tttcaacacc
25801 aaaactcaaa gaagactacc aaattggtgg ttattctgag gataggcact caggtgttaa
25861 agactatgtc gttgtacatg gctatttcac cgaagtttac taccagcttg agtctacaca
25921 aattactaca gacactggta ttgaaaatgc tacattcttc atctttaaca agcttgttaa
25981 agacccaccg aatgtgcaaa tacacacaat cgacggctct tcaggagttg ctaatccagc
26041 aatggatcca atttatgatg agccgacgac gactactagc gtgcctttgt aagcacaaga

Gene E underscored→

26101 aagtgagtac gaacttATGt actcattcgt ttcggaagaa acaggtacgt taatagttaa
26161 tagcgtactt ctttttcttg ctttcgtggt attcttgcta gtcacactag ccatccttac
26221 tgcgcttcga ttgtgtgcgt actgctgcaa tattgttaac gtgagtttag taaaaccaac
26281 ggtttacgtc tactcgcgtg ttaaaaatct gaactcttct gaaggagttc ctgatcttct
26341 ggtc*TAA*acg aactaactat tattattatt ctgtttggaa ctttaacatt gcttatcATG

<-Gene M underscored→

26401 gcagacaacg gtactattac cgttgaggag cttaaacaac tcctggaaca atggaaccta
26461 gtaataggtt tcctattcct agcctggatt atgttactac aatttgccta ttctaatcgg
26521 aacaggtttt tgtacataat aaagcttgtt ttcctctggc tcttgtggcc agtaacactt
26581 gcttgttttg tgcttgctgc tgtctacaga attaattggg tgactggcgg gattgcgatt
26641 gcaatggctt gtattgtagg cttgatgtgg cttagctact tcgttgcttc cttcaggctg

26701 tttgctcgta cccgctcaat gtggtcattc aacccagaaa caaacattct tctcaatgtg
26761 cctctccggg ggacaattgt gaccagaccg ctcatggaaa gtgaacttgt cattggtgct
26821 gtgatcattc gtggtcactt gcgaatggcc ggacactccc tagggcgctg tgacattaag
26881 gacctgccaa aagagatcac tgtggctaca tcacgaacgc tttcttatta caaattagga
26941 gcgtcgcagc gtgtaggcac tgattcaggt tttgctgcat acaaccgcta ccgtattgga
27001 aactataaat taaatacaga ccacgccggt agcaacgaca atattgcttt gctagtacag
27061 *TAA*gtgacaa cagatgtttc atcttgttga cttccaggtt acaatagcag agatattgat
27121 tatcattatg aggactttca ggattgctat ttggaatctt gacgttataa taagttcaat
27181 agtgagacaa ttatttaagc ctctaactaa gaagaattat tcggagttag atgatgaaga
27241 acctatggag ttagattatc cataaaacga acatgaaaat tattctcttc ctgacattga
27301 ttgtatttac atcttgcgag ctatatcact atcaggagtg tgttagaggt acgactgtac
27361 tactaaaaga accttgccca tcaggaacat acgagggcaa ttcaccattt caccctcttg
27421 ctgacaataa atttgcacta acttgcacta gcacacactt tgcttttgct tgtgctgacg
27481 gtactcgaca tacctatcag ctgcgtgcaa gatcagtttc accaaaactt ttcatcagac
27541 aagaggaggt tcaacaagag ctctactcgc cactttttct cattgttgct gctctagtat
27601 ttttaatact ttgcttcacc attaagagaa agacagaatg aatgagctca ctttaattga
27661 cttctatttg tgctttttag cctttctgct attccttgtt ttaataatgc ttattatatt
27721 ttggttttca ctcgaaatcc aggatctaga agaaccttgt accaaagtct aaacgaacat
27781 gaaacttctc attgttttga cttgtatttc tctatgcagt tgcatatgca ctgtagtaca
27841 gcgctgtgca tctaataaac ctcatgtgct tgaagatcct tgtaaggtac aacactaggg
27901 gtaatactta tagcactgct tggctttgtg ctctaggaaa ggttttacct tttcatagat
27961 ggcacactat ggttcaaaca tgcacaccta atgttactat caactgtcaa gatccagctg
28021 gtggtgcgct tatagctagg tgttggtacc ttcatgaagg tcaccaaact gctgcattta

<-Gene N underscored->

28081 gagacgtact tgttgtttta aataaacgaa caaattaaaA TGtctgataa tggaccccaa
28141 tcaaaccaac gtagtgcccc ccgcattaca tttggtggac ccacagattc aactgacaat
28201 aaccagaatg gaggacgcaa tggggcaagg ccaaaacagc gccgacccca aggtttaccc
28261 aataatactg cgtcttggtt cacagctctc actcagcatg gcaaggagga acttagattc
28321 cctcgaggcc agggcgttcc aatcaacacc aatagtggtc cagatgacca aattggctac
28381 taccgaagag ctacccgacg agttcgtggt ggtgacggca aaatgaaaga gctcagcccc
28441 agatggtact tctattacct aggaactggc ccagaagctt cacttcccta cggcgctaac
28501 aaagaaggca tcgtatgggt tgcaactgag ggagccttga atacacccaa agaccacatt
28561 ggcacccgca atcctaataa caatgctgcc accgtgctac aacttcctca aggaacaaca
28621 ttgccaaaag gcttctacgc agagggaagc agaggcggca gtcaagcctc ttctcgctcc
28681 tcatcacgta gtcgcggtaa ttcaagaaat tcaactcctg gcagcagtag gggaaattct
28741 cctgctcgaa tggctagcgg aggtggtgaa actgccctcg cgctattgct gctagacaga
28801 ttgaaccagc ttgagagcaa agtttctggt aaaggccaac aacaacaagg ccaaactgtc
28861 actaagaaat ctgctgctga ggcatctaaa aagcctcgcc aaaaacgtac tgccacaaaa
28921 cagtacaacg tcactcaagc atttgggaga cgtggtccag aacaaaccca aggaaatttc
28981 ggggaccaag acctaatcag acaaggaact gattacaaac attggccgca aattgcacaa
29041 tttgctccaa gtgcctctgc attctttgga atgtcacgca ttggcatgga agtcacacct
29101 tcgggaacat ggctgactta tcatggagcc attaaattgg atgacaaaga tccacaattc
29161 aaagacaacg tcatactgct gaacaagcac attgacgcat acaaaacatt cccaccaaca
29221 gagcctaaaa aggacaaaaa gaaaaagact gatgaagctc agcctttgcc gcagagacaa
29281 aagaagcagc ccactgtgac tcttcttcct gcggctgaca tggatgattt ctccagacaa
29341 cttcaaaatt ccatgagtgg agcttctgct gattcaactc aggca*TAAac actcatgatg*

<-3'UTR ->

29401 accacacaag gcagatgggc tatgtaaacg ttttcgcaat tccgtttacg atacatagtc
29461 tactcttgtg cagaatgaat tctcgtaact aaacagcaca agtaggttta gttaacttta
29521 atctcacata gcaatcttta atcaatgtgt aacattaggg aggacttgaa agagccacca
29581 cattttcatc gaggccacgc ggagtacgat cgagggtaca gtgaataatg ctagggagag
29641 ctgcctatat ggaagagccc taatgtgtaa aattaatttt agtagtgcta tccccatgtg
*29701 attttaatag cttcttagga gaatgacaaa aaaaaaaaaa aaaaaaaaaa a**

The following subsequences are shown and annotated above by underscoring the coding sequences of interest with the initiation codon ATG in uppercase characters, and the stop codon in uppercase italic characters.

The individual coding sequences and translated amino acid sequences are provided below:

1. The coding sequence for the S (spike) protein, SEQ ID NO:4, is from nt 21492 to 25259 of SEQ ID NO:3, which comprises *3768 nt* that encode *1255 residues + stop codon*

As established by Krokhin et al. (2003), the glycosylated spike protein (as well as the nucleocapsid protein) can be detected in infected cell culture supernatants with antisera from SARS patients

SEQ ID NO:4

ATG ttt att ttc tta tta ttt ctt act ctc act agt ggt agt gac ctt gac cgg tgc
acc act ttt gat gat gtt caa gct cct aat tac act caa cat act tca tct atg agg
ggg gtt tac tat cct gat gaa att ttt aga tca gac act ctt tat tta act cag gat
tta ttt ctt cca ttt tat tct aat gtt aca ggg ttt cat act att aat cat acg ttt
ggc aac cct gtc ata cct ttt aag gat ggt att tat ttt gct gcc aca gag aaa tca
aat gtt gtc cgt ggt tgg gtt ttt ggt tct acc atg aac aac aag tca cag tcg gtg
att att att aac aat tct act aat gtt gtt ata cga gca tgt aac ttt gaa ttg tgt
gac aac cct ttc ttt gct gtt tct aaa ccc atg ggt aca cag aca cat act atg ata
ttc gat aat gca ttt aat tgc act ttc gag tac ata tct gat gcc ttt tcg ctt gat
gtt tca gaa aag tca ggt aat ttt aaa cac tta cga gag ttt gtg ttt aaa aat aaa
gat ggg ttt ctc tat gtt tat aag ggc tat caa cct ata gat gta gtt cgt gat cta
cct tct ggt ttt aac act ttg aaa cct att ttt aag ttg cct ctt ggt att aac att
aca aat ttt aga gcc att ctt aca gcc ttt tca cct gct caa gac att tgg ggc acg
tca gct gca gcc tat ttt gtt ggc tat tta aag cca act aca ttt atg ctc aag tat
gat gaa aat ggt aca atc aca gat gct gtt gat tgt tct caa aat cca ctt gct gaa
ctc aaa tgc tct gtt aag agc ttt gag att gac aaa gga att tac cag acc tct aat
ttc agg gtt gtt ccc tca gga gat gtt gtg aga ttc cct aat att aca aac ttg tgt
cct ttt gga gag gtt ttt aat gct act aaa ttc cct tct gtc tat gca tgg gag aga
aaa aaa att tct aat tgt gtt gct gat tac tct gtg ctc tac aac tca aca ttt ttt
tca acc ttt aag tgc tat ggc gtt tct gcc act aag ttg aat gat ctt tgc ttc tcc
aat gtc tat gca gat tct ttt gta gtc aag gga gat gat gta aga caa ata gcg cca
gga caa act ggt gtt att gct gat tat aat tat aaa ttg cca gat gat ttc atg ggt
tgt gtc ctt gct tgg aat act agg aac att gat gct act tca act ggt aat tat aat
tat aaa tat agg tat ctt aga cat ggc aag ctt agg ccc ttt gag aga gac ata tct
aat gtg cct ttc tcc cct gat ggc aaa cct tgc acc cca cct gct ctt aat tgt tat
tgg cca tta aat gat tat ggt ttt tac acc act act ggc att ggc tac caa cct tac
aga gtt gta gta ctt tct ttt gaa ctt tta aat gca ccg gcc acg gtt tgt gga cca
aaa tta tcc act gac ctt att aag aac cag tgt gtc aat ttt aat ttt aat gga ctc
act ggt act ggt gtg tta act cct tct tca aag aga ttt caa cca ttt caa caa ttt
ggc cgt gat gtt tct gat ttc act gat tcc gtt cga gat cct aaa aca tct gaa ata
tta gac att tca cct tgc gct ttt ggg ggt gta agt gta att aca cct gga aca aat
gct tca tct gaa gtt gct gtt cta tat caa gat gtt aac tgc act gat gtt tct aca
gca att cat gca gat caa ctc aca cca gct tgg cgc ata tat tct act gga aac aat
gta ttc cag act caa gca ggc tgt ctt ata gga gct gag cat gtc gac act tct tat
gag tgc gac att cct att gga gct ggc att tgt gct agt tac cat aca gtt tct tta
tta cgt agt act agc caa aaa tct att gtg gct tat act atg tct tta ggt gct gat
agt tca att gct tac tct aat aac acc att gct ata cct act aac ttt tca att agc
att act aca gaa gta atg cct gtt tct atg gct aaa acc tcc gta gat tgt aat atg
tac atc tgc gga gat tct act gaa tgt gct aat ttg ctt ctc caa tat ggt agc ttt
tgc aca caa cta aat cgt gca ctc tca ggt att gct gct gaa cag gat cgc aac aca
cgt gaa gtg ttc gct caa gtc aaa caa atg tac aaa acc cca act ttg aaa tat ttt
ggt ggt ttt aat ttt tca caa ata tta cct gac cct cta aag cca act aag agg tct
ttt att gag gac ttg ctc ttt aat aag gtg aca ctc gct gat gct ggc ttc atg aag
caa tat ggc gaa tgc cta ggt gat att aat gct aga gat ctc att tgt gcg cag aag
ttc aat gga ctt aca gtg ttg cca cct ctg ctc act gat gat atg att gct gcc tac
act gct gct cta gtt agt ggt act gcc act gct gga tgg aca ttt ggt gct ggc gct
gct ctt caa ata cct ttt gct atg caa atg gca tat agg ttc aat ggc att gga gtt
acc caa aat gtt ctc tat gag aac caa aaa caa atc gcc aac caa ttt aac aag gcg
att agt caa att caa gaa tca ctt aca aca aca tca act gca ttg ggc aag ctg caa
gac gtt gtt aac cag aat gct caa gca tta aac aca ctt gtt aaa caa ctt agc tct
aat ttt ggt gca att tca agt gtg cta aat gat atc ctt tcg cga ctt gat aaa gtc
gag gcg gag gta caa att gac agg tta att aca ggc aga ctt caa agc ctt caa acc
tat gta aca caa caa cta atc agg gct gct gaa atc agg gct tct gct aat ctt gct
gct act aaa atg tct gag tgt gtt ctt gga caa tca aaa aga gtt gac ttt tgt gga

```
aag ggc tac cac ctt atg tcc ttc cca caa gca gcc ccg cat ggt gtt gtc ttc cta
cat gtc acg tat gtg cca tcc cag gag agg aac ttc acc aca gcg cca gca att tgt
cat gaa ggc aaa gca tac ttc cct cgt gaa ggt gtt ttt gtg ttt aat ggc act tct
tgg ttt att aca cag agg aac ttc ttt tct cca caa ata att act aca gac aat aca
ttt gtc tca gga aat tgt gat gtc gtt att ggc atc att aac aac aca gtt tat gat
cct ctg caa cct gag ctt gac tca ttc aaa gaa gag ctg gac aag tac ttc aaa aat
cat aca tca cca gat gtt gat ctt ggc gac att tca ggc att aac gct tct gtc gtc
aac att caa aaa gaa att gac cgc ctc aat gag gtc gct aaa aat tta aat gaa tca
ctc att gac ctt caa gaa ttg gga aaa tat gag caa tat att aaa tgg cct tgg tat
gtt tgg ctc ggc ttc att gct gga cta att gcc atc gtc atg gtt aca atc ttg ctt
tgt tgc atg act agt tgt tgc agt tgc ctc aag ggt gca tgc tct tgt ggt tct tgc
tgc aag ttt gat gag gat gac tct gag cca gtt ctc aag ggt gtc aaa tta cat tac
aca TAA
```

Glycosylation sites of this protein include residues encoded by codons at the following positions: 21843-21845; 21846-21848; 22170-22172; 22296-22298; and 23838-23840.

The encoded amino acid sequence of the S polypeptide (SEQ ID NO:5) is:

```
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL   60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS  120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK  180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP  240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY  300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF  360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV  420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND  480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP  540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCAFGG VSVITPGTNA SSEVAVLYQD  600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY  660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC  720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG  780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL  840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE  900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN  960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK 1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN 1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN 1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL 1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT      1255
```

2. The coding sequence for the E (envelope, or "small envelope") protein (SEQ ID NO:6) is from nt 26117 to 26347 of SEQ ID NO:3, which comprises 231 nt that encode 76 aa's + stop codon

SEQ ID NO:6

```
ATG tac tca ttc gtt tcg gaa gaa aca ggt acg tta ata gtt aat agc gta ctt ctt
ttt ctt gct ttc gtg gta ttc ttg cta gtc aca cta gcc atc ctt act gcg ctt cga
ttg tgt gcg tac tgc tgc aat att gtt aac gtg agt tta gta aaa cca acg gtt tac
gtc tac tcg cgt gtt aaa aat ctg aac tct tct gaa gga gtt cct gat ctt ctg gtc
TAA
```

The encoded amino acid sequence of the E polypeptide (SEQ ID NO:7) is:

```
MYSFVSEETG TLIVNSVLLF LAFVVFLLVT LAILTALRLC AYCCNIVNVS LVKPTVYVYS  60
RVKNLNSSEG VPDLLV                                                76
```

3. The coding sequence for the M (membrane protein (SEQ ID NO:8) is from nt 26348 to 26353 of SEQ ID NO:3, which comprises 666 nt encoding 221 aa + stop codon

SEQ ID NO:8

```
ATG gca gac aac ggt act att acc gtt gag gag ctt aaa caa ctc ctg gaa caa tgg
aac cta gta ata ggt ttc cta ttc cta gcc tgg att atg tta cta caa ttt gcc tat
tct aat cgg aac agg ttt ttg tac ata ata aag ctt gtt ttc ctc tgg ctc ttg tgg
cca gta aca ctt gct tgt ttt gtg ctt gct gct gtc tac aga att aat tgg gtg act
ggc ggg att gcg att gca atg gct tgt att gta ggc ttg atg tgg ctt agc tac ttc
gtt gct tcc ttc agg ctg ttt gct cgt acc cgc tca atg tgg tca ttc aac cca gaa
aca aac att ctt ctc aat gtg cct ctc cgg ggg aca att gtg acc aga ccg ctc atg
gaa agt gaa ctt gtc att ggt gct gtg atc att cgt ggt cac ttg cga atg gcc gga
cac tcc cta ggg cgc tgt gac att aag gac ctg cca aaa gag atc act gtg gct aca
tca cga acg ctt tct tat tac aaa tta gga gcg tcg cag cgt gta ggc act gat tca
ggt ttt gct gca tac aac cgc tac cgt att gga aac tat aaa tta aat aca gac cac
gcc ggt agc aac gac aat att gct ttg cta gta cag TAA
```

The encoded amino acid sequence of the M polypeptide (SEQ ID NO:9) is:

```
MADNGTITVE ELKQLLEQWN LVIGFLFLAW IMLLQFAYSN RNRFLYIIKL VFLWLLWPVT   60
LACFVLAAVY RINWVTGGIA IAMACIVGLM WLSYFVASFR LFARTRSMWS FNPETNILLN  120
VPLRGTIVTR PLMESELVIG AVIIRGHLRM AGHSLGRCDI KDLPKEITVA TSRTLSYYKL  180
GASQRVGTDS GFAAYNRYRI GNYKLNTDHA GSNDNIALLV Q                      221
```

4. The coding sequence for the N (nucleocapsid protein (SEQ ID NO:10) is from nt 28120 to 29388of SEQ ID NO:3, which comprises 1269 nt encoding 422 aa + stop codon.

SEQ ID NO:10

```
ATG tct gat aat gga ccc caa tca aac caa cgt agt gcc ccc cgc att aca ttt ggt
gga ccc aca gat tca act gac aat aac cag aat gga gga cgc aat ggg gca agg cca
aaa cag cgc cga ccc caa ggt tta ccc aat aat act gcg tct tgg ttc aca gct ctc
act cag cat ggc aag gag gaa ctt aga ttc cct cga ggc cag ggc gtt cca atc aac
acc aat agt ggt cca gat gac caa att ggc tac tac cga aga gct acc cga cga gtt
cgt ggt ggt gac ggc aaa atg aaa gag ctc agc ccc aga tgg tac ttc tat tac cta
gga act ggc cca gaa gct tca ctt ccc tac ggc gct aac aaa gaa ggc atc gta tgg
gtt gca act gag gga gcc ttg aat aca ccc aaa gac cac att ggc acc cgc aat cct
aat aac aat gct gcc acc gtg cta caa ctt cct caa gga aca aca ttg cca aaa ggc
ttc tac gca gag gga agc aga ggc ggc agt caa gcc tct tct cgc tcc tca tca cgt
agt cgc ggt aat tca aga aat tca act cct ggc agc agt agg gga aat tct cct gct
cga atg gct agc gga ggt ggt gaa act gcc ctc gcg cta ttg ctg cta gac aga ttg
aac cag ctt gag agc aaa gtt tct ggt aaa ggc caacaa caa caa ggc caa act gtc
act aag aaa tct gct gct gag gca tct aaa aag cct cgc caa aaa cgt act gcc aca
aaa cag tac aac gtc act caa gca ttt ggg aga cgt ggt cca gaa caa acc caa gga
aat ttc ggg gac caa gac cta atc aga caa gga act gat tac aaa cat tgg ccg caa
att gca caa ttt gct cca agt gcc tct gca ttc ttt gga atg tca cgc att ggc atg
gaa gtc aca cct tcg gga aca tgg ctg act tat cat gga gcc att aaa ttg gat gac
aaa gat cca caa ttc aaa gac aac gtc ata ctg ctg aac aag cac att gac gca tac
aaa aca ttc cca cca aca gag cct aaa aag gac aaa aag aaa aag act gat gaa gct
cag cct ttg ccg cag aga caa aag aag cag ccc act gtg act ctt ctt cct gcg gct
gac atg gat gat ttc tcc aga caa ctt caa aat tcc atg agt gga gct tct gct gat
tca act cag gca TAA
```

The encoded amino acid sequence of the E polypeptide (SEQ ID NO:11) is:

```
MSDNGPQSNQ RSAPRITFGG PTDSTDNNQN GGRNGARPKQ RRPQGLPNNT ASWFTALTQH   60
GKEELRFPRG QGVPINTNSG PDDQIGYYRR ATRRVRGGDG KMKELSPRWY FYYLGTGPEA  120
SLPYGANKEG IVWVATEGAL NTPKDHIGTR NPNNNAATVL QLPQGTTLPK GFYAEGSRGG  180
SQASSRSSSR SRGNSRNSTP GSSRGNSPAR MASGGGETAL ALLLLDRLNQ LESKVSGKGQ  240
QQQGQTVTKK SAAEASKKPR QKRTATKQYN VTQAFGRRGP EQTQGNFGDQ DLIRQGTDYK  300
```

HWPQIAQFAP SASAFFGMSR IGMEVTPSGT WLTYHGAIKL DDKDPQFKDN VILLNKHIDA 360
YKTFPPTEPK KDKKKKTDEA QPLPQRQKKQ PTVTLLPAAD MDDFSRQLQN SMSGASADST 420
QA 422

As established by Krokhin, O. *et al.*, 2003, *Mol Cell Proteomics* 2:346-56, the N-terminal methionine (encoded by the initiation ATG codon, is removed in the virion protein when it is processed, and all other methionines are oxidized, and the resulting N-terminal serine is acetylated.

CLONING OF THE GENOME OF THE TW1 STRAIN OF SARS-CoV

The presently exemplified and preferred sequences are based on the Taiwanese strain, TW1, of SARS-CoV. The SuperScript cDNA system (Invitrogen, Carlsbad, CA, USA) was used to reverse transcribe the RNA template into cDNA (Hsueh, PR *et al.*,. *Emerg Infect Dis*, *9:* 1163-1167, 2003). To sequence the viral genome, 25 primer sets were designed based on the cDNA sequence data from the Tor2 SARS isolate (accession no. NC_004718, *supra*). See Figure 19 and Table 1. After PCR amplification, products were analyzed by agarose gel electrophoresis and then processed for direct sequencing reactions. Sequences were assembled and edited to obtain the sequence of the genome of the TW1 strain of SARS-CoV, which was subsequently deposited in GenBank (as accession number AY291451; available at WWW URL ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=30698326).

data from the Tor2 SARS isolate (accession no. NC_004718, *supra*). See Figure 19 and Table 1. After PCR amplification, products were analyzed by agarose gel electrophoresis and then processed for direct sequencing reactions. Sequences were assembled and edited to obtain the sequence of the genome of the TW1 strain of SARS-CoV, which was subsequently deposited in GenBank (as accession number AY291451; available at WWW URL

Table 1. Summary of the 25 overlapping SARS-CoV TW-1 isolate cDNA clones sequenced and available. The cDNA sections are in the vector, between the BamHI and EcoRI cloning sites. Forward and reverse sequencing primers are shown.

Clone	SARS Nucleotides	Forward Sequencing Primer	SEQ ID NO:	Reverse Sequencing Primer	SEQ ID NO:
1	1-1471	CTACCCAGGAAAAGCCAACC	52	CAACATAGGCAAACACACAGC	53
2	1345-2675	GAAGGACCTACTACATGTGGG	54	CTTCCCAAACAGTATCTTCTCC	55
3	2519-3918	CAAGGAGCAGCTGCAACTAC	56	TGTTCTGAGAATCATGGTAAAG C	57
4	3757-5131	GTCTTTACAAGTGTGCGTGCAG	58	GCCTCTTGAAGTGCTGGTGC	59
5	4967-6344	TGACATATGGACAGCAGTTTGG	60	TTCGGTAGTTTTCACGTCACAC	61
6	6166-7577	TTGAATGGCGATGTAGTGGC	62	CTGGTCAGTAGGGTTGATTGG	63
7	7395-8788	CCCGTTTCTGCAATGGTTAGG	64	GCTCTCAGCACAGTACCCGG	65
8	8603-10023	GCCAGTACATACATTGTCAATCC	66	TCCATTAAGAGTTGTAGTTCCA C	67
9	9835-11198	GCGTAGCGAGACACTGTTGCC	68	CATCATCATAAACAGTGCGAGC	69
10	11017-12421	GTTCAAAGTACACAGTGGTCAC	70	TCAACAACTTGCTGGATTTCCC	71
11	12250-13658	GACCCAAATGTACAAACAGGC	72	CTGACGTGATATATGTGGTACC	73
12	13451-14834	GGCACTAGTACTGATGTCGTC	74	GATGACATTACGCTTAGTATAC G	75
13	14672-16052	CTTTTCAAACTGTCAAACCCGG	76	AGCCTGCAAGACTGTATGTGG	77
14	15859-17253	TTACGTGTACCTGCCTTACCC	78	AGTCATAATTAGTAGCCATAGA G	79
15	17054-18445	CGGACTTGCTCTCTATTACCC	80	CACGACTCTGTCTGACAATCC	81
16	18276-19658	CAACTAGAGATGCTGTGGGTAC	82	GCTCAAATGCAACATTAACAGG	84
17	19450-20845	CCATGCAAATGAGTACCGACAG	84	CTGAATCGACAAGTAGTGTGC C	85
18	20683-22072	AAGTGTGACCTTCAGAATTATGG	86	ACCAGAAGGTAGATCACGAAC	88
19	21871-23223	ACTAATGTTGTTATACGAGCATG	88	CAGATGAAGCATTTGTTCCAGG	90
20	23061-24439	ATCCACTGACCTTATTAAGAACC	90	AGCAGAAGCCCTGATTTCAGC AGC	92
21	24260-25666	CAACAACATCAACTGCATTGGG	92	TCATAGTTATGTGTGTGCCAGC	94
22	25474-26868	CAATAAAAGATGGCAGCTAGC	94	GTAGCCACAGTGATCTCTTTTC	96

data from the Tor2 SARS isolate (accession no. NC_004718, *supra*). See Figure 19 and Table 1. After PCR amplification, products were analyzed by agarose gel electrophoresis and then processed for direct sequencing reactions. Sequences were assembled and edited to obtain the sequence of the genome of the TW1 strain of SARS-CoV, which was subsequently deposited in GenBank (as accession number AY291451; available at WWW URL ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=30698326). This data is based on Yeh, S-H *et al.*, *Proc. Natl. Acad. Sci. U.S.A. 101*:2542-2547 (2004) and later deposits by the same group (see URL).

The genomic sequence of the TW1 strain, nt 1-29729 is shown below (SEQ ID NO:12) Annotation is as in SEQ ID NO:3 above (the TOR2 strain)

SEQ ID NO:12

```
  1 atattaggtt tttacctacc caggaaaagc caaccaacct cgatctcttg tagatctgtt
 61 ctctaaacga actttaaaat ctgtgtagct gtcgctcggc tgcatgccta gtgcacctac
121 gcagtataaa caataataaa ttttactgtc gttgacaaga aacgagtaac tcgtccctct
181 tctgcagact gcttacggtt tcgtccgtgt tgcagtcgat catcagcata cctaggtttc
241 gtccgggtgt gaccgaaagg taagatggag agccttgttc ttggtgtcaa cgagaaaaca
```

```
 301 cacgtccaac tcagtttgcc tgtccttcag gttagagacg tgctagtgcg tggcttcggg
 361 gactctgtgg aagaggccct atcggaggca cgtgaacacc tcaaaaatgg cacttgtggt
 421 ctagtagagc tggaaaaagg cgtactgccc cagcttgaac agccctatgt gttcattaaa
 481 cgttctgatg ccttaagcac caatcacggc cacaaggtcg ttgagctggt tgcagaaatg
 541 gacggcattc agtacggtcg tagcggtata acactgggag tactcgtgcc acatgtgggc
 601 gaaaccccaa ttgcataccg caatgttctt cttcgtaaga acggtaataa gggagccggt
 661 ggtcatagct atggcatcga tctaaagtct tatgacttag gtgacgagct tggcactgat
 721 cccattgaag attatgaaca aaactggaac actaagcatg gcagtggtgc actccgtgaa
 781 ctcactcgtg agctcaatgg aggtgcagtc actcgctatg tcgacaacaa tttctgtggc
 841 ccagatgggt accctcttga ttgcatcaaa gattttctcg cacgcgcggg caagtcaatg
 901 tgcactcttt ccgaacaact tgattacatc gagtcgaaga gaggtgtcta ctgctgccgt
 961 gaccatgagc atgaaattgc ctggttcact gagcgctctg ataagagcta cgagcaccag
1021 acacccttcg aaattaagag tgccaagaaa tttgacactt tcaaagggga atgcccaaag
1081 tttgtgtttc ctcttaactc aaaagtcaaa gtcattcaac cacgtgttga aaagaaaaag
1141 actgagggtt tcatggggcg tatacgctct gtgtaccctg ttgcatctcc acaggagtgt
1201 aacaatatgc acttgtctac cttgatgaaa tgtaatcatt gcgatgaagt ttcatggcag
1261 acgtgcgact ttctgaaagc cacttgtgaa cattgtggca ctgaaaattt agttattgaa
1321 ggacctacta catgtgggta cctacctact aatgctgtag tgaaaatgcc atgtcctgcc
1381 tgtcaagacc cagagattgg acctgagcat agtgttgcag attatcacaa ccactcaaac
1441 attgaaactc gactccgcaa gggaggtagg actagatgtt ttggaggctg tgtgtttgcc
1501 tatgttggct gctataataa gcgtgcctac tgggttcctc gtgctagtgc tgatattggc
1561 tcaggccata ctggcattac tggtgacaat gtggagacct tgaatgagga tctccttgag
1621 atactgagtc gtgaacgtgt taacattaac attgttggcg attttcattt gaatgaagag
1681 gttgccatca ttttggcatc tttctctgct tctacaagtg cctttattga cactataaag
1741 agtcttgatt acaagtcttt caaaaccatt gttgagtcct gcggtaacta taaagttacc
1801 aagggaaagc ccgtaaaagg tgcttggaac attggacaac agagatcagt tttaacacca
1861 ctgtgtggtt ttccctcaca ggctgctggt gttatcagat caatttttgc gcgcacactt
1921 gatgcagcaa accactcaat tcctgatttg caaagagcag ctgtcaccat acttgatggt
1981 atttctgaac agtcattacg tcttgtcgac gccatggttt atacttcaga cctgctcacc
2041 aacagtgtca ttattatggc atatgtaact ggtggtcttg tacaacagac ttctcagtgg
2101 ttgtctaatc ttttgggcac tactgttgaa aaactcaggc ctatctttga atggattgag
2161 gcgaaactta gtgcaggagt tgaatttctc aaggatgctt gggagattct caaatttctc
2221 attacaggtg tttttgacat cgtcaagggt caaatacagg ttgcttcaga taacatcaag
2281 gattgtgtaa aatgcttcat tgatgttgtt aacaaggcac tcgaaatgtg cattgatcaa
2341 gtcactatcg ctggcgcaaa gttgcgatca ctcaacttag gtgaagtctt catcgctcaa
2401 agcaagggac tttaccgtca gtgtatacgt ggcaaggagc agctgcaact actcatgcct
2461 cttaaggcac caaaagaagt aacctttctt gaaggtgatt cacatgacac agtacttacc
2521 tctgaggagg ttgttctcaa gaacggtgaa ctcgaagcac tcgagacgcc cgttgatagc
2581 ttcacaaatg gagctatcgt tggcacacca gtctgtgtaa atggcctcat gctcttagag
2641 attaaggaca aagaacaata ctgcgcattg tctcctggtt tactggctac aaacaatgtc
2701 tttcgcttaa aggggggtgc accaattaaa ggtgtaacct ttggagaaga tactgtttgg
2761 gaagttcaag gttacaagaa tgtgagaatc acatttgagc ttgatgaacg tgttgacaaa
2821 gtgcttaatg aaaagtgctc tgtctacact gttgaatccg gtaccgaagt tactgagttt
2881 gcatgtgttg tagcagaggc tgttgtgaag actttacaac cagtttctga tctccttacc
2941 aacatgggta ttgatcttga tgagtggagt gtagctacat tctacttatt tgatgatgct
3001 ggtgaagaaa acttttcatc acgtatgtat tgttcctttt accctccaga tgaggaagaa
3061 gaggacgatg cagagtgtga ggaagaagaa attgatgaaa cctgtgaaca tgagtacggt
3121 acagaggatg attatcaagg tctccctctg gaatttggtg cctcggctga aacagttcga
3181 gttgaggaag aagaagagga agactggctg gatgatacta ctgagcaatc agagattgag
3241 ccagaaccag aacctacacc tgaagaacca gttaatcagt ttactggtta tttaaaactt
3301 actgacaatg ttgccattaa atgtgttgac atcgttaagg aggcacaaag tgctaatcct
3361 atggtgattg taaatgctgc taacatacac ctgaaacatg gtggtggtgt agcaggtgca
3421 ctcaacaagg caaccaatgg tgccatgcaa aaggagagtg atgattacat taagctaaat
3481 ggccctctta cagtaggagg gtcttgtttg ctttctggac ataatcttgc taagaagtgt
3541 ctgcatgttg ttggacctaa cctaaatgca ggtgaggaca tccagcttct taaggcagca
3601 tatgaaaatt tcaattcaca ggacatctta cttgcaccat tgttgtcagc aggcatattt
3661 ggtgctaaac cacttcagtc tttacaagtg tgcgtgcaga cggttcgtac acaggtttat
3721 attgcagtca atgacaaagc tctttatgag caggttgtca tggattatct tgataacctg
3781 aagcctagag tggaagcacc taaacaagag gagccaccaa acacagaaga ttccaaaact
3841 gaggagaaat ctgtcgtaca gaagcctgtc gatgtgaagc caaaaattaa ggcctgcatt
3901 gatgaggtta ccacaacact ggaagaaact aagtttctta ccaataagtt actcttgttt
3961 gctgatatca atggtaagct ttaccatgat tctcagaaca tgcttagagg tgaagatatg
4021 tctttccttg agaaggatgc accttacatg gtaggtgatg ttatcactag tggtgatatc
4081 acttgtgttg taataccctc caaaaaggct ggtggcacta ctgagatgct ctcaagagct
4141 ttgaagaaag tgccagttga tgagtatata accacgtacc ctggacaagg atgtgctggt
4201 tatacacttg aggaagctaa gactgctctt aagaaatgca aatctgcatt ttatgtacta
```

4261 ccttcagaag cacctaatgc taaggaagag attctaggaa ctgtatcctg gaatttgaga
4321 gaaatgcttg ctcatgctga agagacaaga aaattaatgc ctatatgcat ggatgttaga
4381 gccataatgg caaccatcca acgtaagtat aaaggaatta aaattcaaga gggcatcgtt
4441 gactatggtg tccgattctt cttttatact agtaaagagc ctgtagcttc tattattacg
4501 aagctgaact ctctaaatga gccgcttgtc acaatgccaa ttggttatgt gacacatggt
4561 tttaatcttg aagaggctgc gcgctgtatg cgttctctta aagctcctgc cgtagtgtca
4621 gtatcatcac cagatgctgt tactacatat aatggatacc tcacttcgtc atcaaagaca
4681 tctgaggagc actttgtaga aacagtttct ttggctggct cttacagaga ttggtcctat
4741 tcaggacagc gtacagagtt aggtgttgaa tttcttaagc gtggtgacaa aattgtgtac
4801 cacactctgg agagccccgt cgagtttcat cttgacggtg aggttctttc acttgacaaa
4861 ctaaagagtc tcttatccct gcgggaggtt aagactataa aagtgttcac aactgtggac
4921 aacactaatc tccacacaca gcttgtggat atgtctatga catatggaca gcagtttggt
4981 ccaacatact tggatggtgc tgatgttaca aaaattaaac ctcatgtaaa tcatgagggt
5041 aagactttct ttgtactacc tagtgatgac acactacgta gtgaagcttt cgagtactac
5101 catactcttg atgagagttt tcttggtagg tacatgtctg ctttaaacca cacaaagaaa
5161 tggaaatttc ctcaagttgg tggtttaact tcaattaaat gggctgataa caattgttat
5221 ttgtctagtg ttttattagc acttcaacag cttgaagtca aattcaatgc accagcactt
5281 caagaggctt attatagagc ccgtgctggt gatgctgcta acttttgtgc actcatactc
5341 gcttacagta ataaaactgt tggcgagctt ggtgatgtca gagaaactat gacccatctt
5401 ctacagcatg ctaatttgga atctgcaaag cgagttctta atgtggtgtg taaacattgt
5461 ggtcagaaaa ctactacctt aacgggtgta gaagctgtga tgtatatggg tactctatct
5521 tatgataatc ttaagacagg tgtttccatt ccatgtgtgt gtggtcgtga tgctacacaa
5581 tatctagtac aacaagagtc ttcttttgtt atgatgtctg caccacctgc tgagtataaa
5641 ttacagcaag gtacattctt atgtgcgaat gagtacactg gtaactatca gtgtggtcat
5701 tacactcata taactgctaa ggagaccctc tatcgtattg acggagctca ccttacaaag
5761 atgtcagagt acaaaggacc agtgactgat gttttctaca aggaaacatc ttacactaca
5821 accatcaagc ctgtgtcgta taaactcgat ggagttactt acacagagat tgaaccaaaa
5881 ttggatgggt attataaaaa ggataatgct tactatacag agcagcctat agaccttgta
5941 ccaactcaac cattaccaaa tgcgagtttt gataatttca aactcacatg ttctaacaca
6001 aaatttgctg atgatttaaa tcaaatgaca ggcttcacaa agccagcttc acgagagcta
6061 tctgtcacat tcttcccaga cttgaatggc gatgtagtgg ctattgacta tagacactat
6121 tcagcgagtt tcaagaaagg tgctaaatta ctgcataagc caattgtttg gcacattaac
6181 caggctacaa ccaagacaac gttcaaacca aacacttggt gtttacgttg tctttggagt
6241 acaaagccag tagatacttc aaattcattt gaagttctgg cagtagaaga cacacaagga
6301 atggacaatc ttgcttgtga aagtcaacaa cccacctctg aagaagtagt ggaaaatcct
6361 accatacaga aggaagtcat agagtgtgac gtgaaaacta ccgaagttgt aggcaatgtc
6421 atacttaaac catcagatga aggtgttaaa gtaacacaag agttaggtca tgaggatctt
6481 atggctgctt atgtggaaaa cacaagcatt accattaaga aacctaatga gctttcacta
6541 gccttaggtt taaaaacaat tgccactcat ggtattgctg caattaatag tgttccttgg
6601 agtaaaattt tggcttatgt caaaccattc ttaggacaag cagcaattac aacatcaaat
6661 tgcgctaaga gattagcaca acgtgtgttt aacaattata tgccttatgt gtttacatta
6721 ttgttccaat tgtgtacttt tactaaaagt accaattcta gaattagagc ttcactacct
6781 acaactattg ctaaaaatag tgttaagagt gttgctaaat tatgtttgga tgccggcatt
6841 aattatgtga agtcacccaa attttctaaa ttgttcacaa tcgctatgtg gctattgttg
6901 ttaagtattt gcttaggttc tctaatctgt gtaactgctg cttttggtgt actcttatct
6961 aattttggtg ctccttctta ttgtaatggc gttagagaat tgtatcttaa ttcgtctaac
7021 gttactacta tggatttctg tgaaggttct tttccttgca gcatttgttt aagtggatta
7081 gactcccttg attcttatcc agctcttgaa accattcagg tgacgatttc atcgtacaag
7141 ctagacttga caattttagg tctggccgct gagtgggttt tggcatatat gttgttcaca
7201 aaattctttt atttattagg tctttcagct ataatgcagg tgttctttgg ctattttgct
7261 agtcatttca tcagcaattc ttggctcatg tggtttatca ttagtattgt acaaatggca
7321 cccgtttctg caatggttag gatgtacatc ttctttgctt ctttctacta catatggaag
7381 agctatgttc atatcatgga tggttgcacc tcttcgactt gcatgatgtg ctataagcgc
7441 aatcgtgcca cacgcgttga gtgtacaact attgttaatg gcatgaagag atctttctat
7501 gtctatgcaa atggaggccg tggcttctgc aagactcaca attggaattg tctcaattgt
7561 gacacatttt gcactggtag tacattcatt agtgatgaag ttgctcgtga tttgtcactc
7621 cagtttaaaa gaccaatcaa ccctactgac cagtcatcgt atattgttga tagtgttgct
7681 gtgaaaaatg gcgcgcttca cctctacttt gacaaggctg gtcaaaagac ctatgagaga
7741 catccgctct cccattttgt caatttagac aatttgagag ctaacaacac taaaggttca
7801 ctgcctatta atgtcatagt ttttgatggc aagtccaaat gcgacgagtc tgcttctaag
7861 tctgcttctg tgtactacag tcagctgatg tgccaaccta ttctgttgct tgaccaagct
7921 cttgtatcag acgttggaga tagtactgaa gtttccgtta agatgtttga tgcttatgtc
7981 gacacctttt cagcaacttt tagtgttcct atggaaaaac ttaaggcact tgttgctaca
8041 gctcacagcg agttagcaaa gggtgtagct ttagatggtg tcctttctac attcgtgtca
8101 gctgcccgac aaggtgttgt tgataccgat gttgacacaa aggatgttat tgaatgtctc
8161 aaactttcac atcactctga cttagaagtg acaggtgaca gttgtaacaa tttcatgctc

8221 acctataata aggttgaaaa catgacgccc agagatcttg gcgcatgtat tgactgtaat
8281 gcaaggcata tcaatgccca agtagcaaaa agtcacaatg tttcactcat ctggaatgta
8341 aaagactaca tgtctttatc tgaacagctg cgtaaacaaa ttcgtagtgc tgccaagaag
8401 aacaacatac cttttagact aacttgtgct acaactagac aggttgtcaa tgtcataact
8461 actaaaatct cactcaaggg tggtaagatt gttagtactt gttttaaact tatgcttaag
8521 gccacattat tgtgcgttct tgctgcattg gtttgttata tcgttatgcc agtacataca
8581 ttgtcaatcc atgatggtta cacaaatgaa atcattggtt acaaagccat tcaggatggt
8641 gtcactcgtg acatcatttc tactgatgat tgttttgcaa ataaacatgc tggttttgac
8701 gcatggttta gccagcgtgg tggttcatac aaaaatgaca aaagctgccc tgtagtagct
8761 gctatcatta caagagagat tggtttcata gtgcctggct taccgggtac tgtgctgaga
8821 gcaatcaatg gtgacttctt gcattttcta cctcgtgttt ttagtgctgt tggcaacatt
8881 tgctacacac cttccaaact cattgagtat agtgattttg ctacctctgc ttgcgttctt
8941 gctgctgagt gtacaatttt taaggatgct atgggcaaac ctgtgccata ttgttatgac
9001 actaatttgc tagagggttc tatttcttat agtgagcttc gtccagacac tcgttatgtg
9061 cttatggatg gttccatcat acagtttcct aacacttacc tggagggttc tgttagagta
9121 gtaacaactt ttgatgctga gtactgtaga catggtacat gcgaaaggtc agaagtaggt
9181 atttgcctat ctaccagtgg tagatgggtt cttaataatg agcattacag agctctatca
9241 ggagttttct gtggtgttga tgcgatgaat ctcatagcta acatctttac tcctcttgtg
9301 caacctgtgg gtgctttaga tgtgtctgct tcagtagtgg ctggtggtat tattgccata
9361 ttggtgactt gtgctgccta ctactttatg aaattcagac gtgtttttgg tgagtacaac
9421 catgttgttg ctgctaatgc acttttgttt ttgatgtctt tcactatact ctgtctggta
9481 ccagcttaca gctttctgcc gggagtctac tcagtctttt acttgtactt gacattctat
9541 ttcaccaatg atgtttcatt cttggctcac cttcaatggt ttgccatgtt ttctcctatt
9601 gtgccttttt ggataacagc aatctatgta ttctgtattt ctctgaagca ctgccattgg
9661 ttctttaaca actatcttag gaaaagagtc atgtttaatg gagttacatt tagtaccttc
9721 gaggaggctg ctttgtgtac ctttttgctc aacaaggaaa tgtacctaaa attgcgtagc
9781 gagacactgt tgccacttac acagtataac aggtatcttg ctctatataa caagtacaag
9841 tatttcagtg gagccttaga tactaccagc tatcgtgaag cagcttgctg ccacttagca
9901 aaggctctaa atgactttag caactcaggt gctgatgttc tctaccaacc accacagaca
9961 tcaatcactt ctgctgttct gcagagtggt tttaggaaaa tggcattccc gtcaggcaaa
10021 gttgaagggt gcatggtaca agtaacctgt ggaactacaa ctcttaatgg attgtggttg
10081 gatgacacag tatactgtcc aagacatgtc atttgcacag cagaagacat gcttaatcct
10141 aactatgaag atctgctcat tcgcaaatcc aaccatagct ttcttgttca ggctggcaat
10201 gttcaacttc gtgttattgg ccattctatg caaaattgtc tgcttaggct taaagttgat
10261 acttctaacc ctaagacacc caagtataaa tttgtccgta tccaacctgg tcaaacattt
10321 tcagttctag catgctacaa tggttcacca tctggtgttt atcagtgtgc catgagacct
10381 aatcatacca ttaaaggttc tttccttaat ggatcatgtg gtagtgttgg ttttaacatt
10441 gattatgatt gcgtgtcttt ctgctatatg catcatatgg agcttccaac aggagtacac
10501 gctggtactg acttagaagg taaattctat ggtccatttg ttgacagaca aactgcacag
10561 gctgcaggta cagacacaac cataacatta aatgttttgg catggctgta tgctgctgtt
10621 atcaatggtg ataggtggtt tcttaataga ttcaccacta ctttgaatga ctttaacctt
10681 gtggcaatga agtacaacta tgaacctttg acacaagatc atgttgacat attgggacct
10741 ctttctgctc aaacaggaat tgccgtctta gatatgtgtg ctgctttgaa agagctgctg
10801 cagaatggta tgaatggtcg tactatcctt ggtagcacta ttttagaaga tgagtttaca
10861 ccatttgatg ttgttagaca atgctctggt gttaccttcc aaggtaagtt caagaaaatt
10921 gttaagggca ctcatcattg gatgctttta actttcttga catcactatt gattcttgtt
10981 caaagtacac agtggtcact gtttttcttt gtttacgaga atgctttctt gccatttact
11041 cttggtatta tggcaattgc tgcatgtgct atgctgcttg ttaagcataa gcacgcattc
11101 ttgtgcttgt ttctgttacc ttctcttgca acagttgctt actttaatat ggtctacatg
11161 cctgctagct gggtgatgcg tatcatgaca tggcttgaat tggctgacac tagcttgtct
11221 ggttataggc ttaaggattg tgttatgtat gcttcagctt tagttttgct tattctcatg
11281 acagctcgca ctgtttatga tgatgctgct agacgtgttt ggacactgat gaatgtcatt
11341 acacttgttt acaaagtcta ctatggtaat gctttagatc aagctatttc catgtgggcc
11401 ttagttattt ctgtaacctc taactattct ggtgtcgtta cgactatcat gtttttagct
11461 agagctatag tgtttgtgtg tgttgagtat tacccattgt tatttattac tggcaacacc
11521 ttacagtgta tcatgcttgt ttattgtttc ttaggctatt gttgctgctg ctactttggc
11581 cttttctgtt tactcaaccg ttacttcagg cttactcttg gtgtttatga ctacttggtc
11641 tctacacaag aatttaggta tatgaactcc caggggcttt tgcctcctaa gagtagtatt
11701 gatgctttca agcttaacat taagttgttg ggtattggag gtaaaccatg tatcaaggtt
11761 gctactgtac agtctaaaat gtctgacgta aagtgcacat ctgtggtact gctctcggtt
11821 cttcaacaac ttagagtaga gtcatcttct aaattgtggg cacaatgtgt acaactccac
11881 aatgatattc ttcttgcaaa agacacaact gaagctttcg agaagatggt ttctcttttg
11941 tctgttttgc tatccatgca gggtgctgta gacattaata ggttgtgcga ggaaatgctc
12001 gataaccgtg ctactcttca ggctattgct tcagaattta gttctttacc atcatatgcc
12061 gcttatgcca ctgcccagga ggcctatgag caggctgtag ctaatggtga ttctgaagtc
12121 gttctcaaaa agttaaagaa atctttgaat gtggctaaat ctgagtttga ccgtgatgct

```
12181 gccatgcaac gcaagttgga aaagatggca gatcaggcta tgacccaaat gtacaaacag
12241 gcaagatctg aggacaagag ggcaaaagta actagtgcta tgcaaacaat gctcttcact
12301 atgcttagga agcttgataa tgatgcactt aacaacatta tcaacaatgc gcgtgatggt
12361 tgtgttccac tcaacatcat accattgact acagcagcca aactcatggt tgttgtccct
12421 gattatggta cctacaagaa cacttgtgat ggtaacacct ttacatatgc atctgcactc
12481 tgggaaatcc agcaagttgt tgatgcggat agcaagattg ttcaacttag tgaaattaac
12541 atggacaatt caccaaattt ggcttggcct cttattgtta cagctctaag agccaactca
12601 gctgttaaac tacagaataa tgaactgagt ccagtagcac tacgacagat gtcctgtgcg
12661 gctggtacca cacaaacagc ttgtactgat gacaatgcac ttgcctacta taacaattcg
12721 aagggaggta ggtttgtgct ggcattacta tcagaccacc aagatctcaa atgggctaga
12781 ttccctaaga gtgatggtac aggtacaatt tacacagaac tggaaccacc ttgtaggttt
12841 gttacagaca caccaaaagg gcctaaagtg aaatacttgt acttcatcaa aggcttaaac
12901 aacctaaata gaggtatggt gctgggcagt ttagctgcta cagtacgtct tcaggctgga
12961 aatgctacag aagtacctgc caattcaact gtgctttcct tctgtgcttt tgcagtagac
13021 cctgctaaag catataagga ttacctagca agtggaggac aaccaatcac caactgtgtg
13081 aagatgttgt gtacacacac tggtacagga caggcaatta ctgtaacacc agaagctaac
13141 atggaccaag agtcctttgg tggtgcttca tgttgtctgt attgtagatg ccacattgac
13201 catccaaatc ctaaaggatt ctgtgacttg aaaggtaagt acgtccaaat acctaccact
13261 tgtgctaatg acccagtggg ttttacactt agaaacacag tctgtaccgt ctgcggaatg
13321 tggaaaggtt atggctgtag ttgtgaccaa ctccgcgaac ccttgatgca gtctgcggat
13381 gcatcaacgt ttttaaacgg gtttgcggtg taagtgcagc ccgtcttaca ccgtgcggca
13441 caggcactag tactgatgtc gtctacaggg cttttgatat ttacaacgaa aaagttgctg
13501 gttttgcaaa gttcctaaaa actaattgct gtcgcttcca ggagaaggat gaggaaggca
13561 atttattaga ctcttacttt gtagttaaga ggcatactat gtctaactac caacatgaag
13621 agactattta taacttggtt aaagattgtc cagcggttgc tgtccatgac tttttcaagt
13681 ttagagtaga tggtgacatg gtaccacata tatcacgtca gcgtctaact aaatacacaa
13741 tggctgattt agtctatgct ctacgtcatt ttgatgaggg taattgtgat acattaaaag
13801 aaatactcgt cacatacaat tgctgtgatg atgattattt caataagaag gattggtatg
13861 acttcgtaga gaatcctgac atcttacgcg tatatgctaa cttaggtgag cgtgtacgcc
13921 aatcattatt aaagactgta caattctgcg atgctatgcg tgatgcaggc attgtaggcg
13981 tactgacatt agataatcag gatcttaatg ggaactggta cgatttcggt gatttcgtac
14041 aagtagcacc aggctgcgga gttcctattg tggattcata ttactcattg ctgatgccca
14101 tcctcacttt gactagggca ttggctgctg agtcccatat ggatgctgat ctcgcaaaac
14161 cacttattaa gtgggatttg ctgaaatatg attttacgga agagagactt tgtctcttcg
14221 accgttattt taaatattgg gaccagacat accatcccaa ttgtattaac tgtttggatg
14281 ataggtgtat ccttcattgt gcaaacttta atgtgttatt ttctactgtg tttccaccta
14341 caagttttgg accactagta agaaaaatat ttgtagatgg tgttcctttt gttgtttcaa
14401 ctggatacca ttttcgtgag ttaggagtcg tacataatca ggatgtaaac ttacatagct
14461 cgcgtctcag tttcaaggaa cttttagtgt atgctgctga tccagctatg catgcagctt
14521 ctggcaattt attgctagat aaacgcacta catgcttttc agtagctgca ctaacaaaca
14581 atgttgcttt tcaaactgtc aaacccggta atttttaataa agacttttat gactttgctg
14641 tgtctaaagg tttctttaag gaaggaagtt ctgttgaact aaaacacttc ttctttgctc
14701 aggatggcaa cgctgctatc agtgattatg actattatcg ttataatctg ccaacaatgt
14761 gtgatatcag acaactccta ttcgtagttg aagttgttga taaatacttt gattgttacg
14821 atggtggctg tattaatgcc aaccaagtaa tcgttaacaa tctggataaa tcagctggtt
14881 tcccatttaa taaatggggt aaggctagac tttattatga ctcaatgagt tatgaggatc
14941 aagatgcact tttcgcgtat actaagcgta atgtcatccc tactataact caaatgaatc
15001 ttaagtatgc cattagtgca aagaatagag ctcgcaccgt agctggtgtc tctatctgta
15061 gtactatgac aaatagacag tttcatcaga aattattgaa gtcaatagcc gccactagag
15121 gagctactgt ggtaattgga acaagcaagt tttacggtgg ctggcataat atgttaaaaa
15181 ctgtttacag tgatgtagaa actccacacc ttatgggttg ggattatcca aaatgtgaca
15241 gagccatgcc taacatgctt aggataatgg cctctcttgt tcttgctcgc aaacataaca
15301 cttgctgtaa cttatcacac cgtttctaca ggttagctaa cgagtgtgcg caagtattaa
15361 gtgagatggt catgtgtggc ggctcactat atgttaaacc aggtggaaca tcatccggtg
15421 atgctacaac tgcttatgct aatagtgtct ttaacatttg tcaagctgtt acagccaatg
15481 taaatgcact tctttcaact gatggtaata agatagctga caagtatgtc cgcaatctac
15541 aacacaggct ctatgagtgt ctctatagaa atagggatgt tgatcatgaa ttcgtggatg
15601 agttttacgc ttacctgcgt aaacatttct ccatgatgat tctttctgat gatgccgttg
15661 tgtgctataa cagtaactat gcggctcaag gtttagtagc tagcattaag aactttaagg
15721 cagttcttta ttatcaaaat aatgtgttca tgtctgaggc aaaatgttgg actgagactg
15781 accttactaa aggacctcac gaattttgct cacagcatac aatgctagtt aaacaaggag
15841 atgattacgt gtacctgcct tacccagatc catcaagaat attaggcgca ggctgttttg
15901 tcgatgatat tgtcaaaaca gatggtacac ttatgattga aaggttcgtg tcactggcta
15961 ttgatgctta cccacttaca aaacatccta atcaggagta tgctgatgtc tttcacttgt
16021 atttacaata cattagaaag ttacatgatg agcttactgg ccacatgttg gacatgtatt
16081 ccgtaatgct aactaatgat aacacctcac ggtactggga acctgagttt tatgaggcta
```

16141 tgtacacacc acatacagtc ttgcaggctg taggtgcttg tgtattgtgc aattcacaga
16201 cttcacttcg ttgcggtgcc tgtattagga gaccattcct atgttgcaag tgctgctatg
16261 accatgtcat ttcaacatca cacaaattag tgttgtctgt taatccctat gtttgcaatg
16321 ccccaggttg tgatgtcact gatgtgacac aactgtatct aggaggtatg agctattatt
16381 gcaagtcaca taagcctccc attagttttc cattatgtgc taatggtcag gtttttggtt
16441 tatacaaaaa cacatgtgta ggcagtgaca atgtcactga cttcaatgcg atagcaacat
16501 gtgattggac taatgctggc gattacatac ttgccaacac ttgtactgag agactcaagc
16561 ttttcgcagc agaaacgctc aaagccactg aggaaacatt taagctgtca tatggtattg
16621 ccactgtacg cgaagtactc tctgacagag aattgcatct ttcatgggag gttggaaaac
16681 ctagaccacc attgaacaga aactatgtct ttactggtta ccgtgtaact aaaaatagta
16741 aagtacagat tggagagtac acctttgaaa aaggtgacta tggtgatgct gttgtgtaca
16801 gaggtactac gacatacaag ttgaatgttg gtgattactt tgtgttgaca tctcacactg
16861 taatgccact tagtgcacct actctagtgc cacaagagca ctatgtgaga attactggct
16921 tgtacccaac actcaacatc tcagatgagt tttctagcaa tgttgcaaat tatcaaaagg
16981 tcggcatgca aaagtactct acactccaag gaccacctgg tactggtaag agtcattttg
17041 ccatcggact tgctctctat tacccatctg ctcgcatagt gtatacggca tgctctcatg
17101 cagctgttga tgccctatgt gaaaaggcat taaaatattt gcccatagat aaatgtagta
17161 gaatcatacc tgcgcgtgcg cgcgtagagt gttttgataa attcaaagtg aattcaacac
17221 tagaacagta tgttttctgc actgtaaatg cattgccaga aacaactgct gacattgtag
17281 tctttgatga aatctctatg gctactaatt atgacttgag tgttgtcaat gctagacttc
17341 gtgcaaaaca ctacgtctat attggcgatc ctgctcaatt accagccccc cgcacattgc
17401 tgactaaagg cacactagaa ccagaatatt ttaattcagt gtgcagactt atgaaaacaa
17461 taggtccaga catgttcctt ggaacttgtc gccgttgtcc tgctgaaatt gttgacactg
17521 tgagtgcttt agtttatgac aataagctaa aagcacacaa ggataagtca gctcaatgct
17581 tcaaaatgtt ctacaaaggt gttattacac atgatgtttc atctgcaatc aacagacctc
17641 aaataggcgt tgtaagagaa tttcttacac gcaatcctgc ttggagaaaa gctgttttta
17701 tctcacctta taattcacag aacgctgtag cttcaaaaat cttaggattg cctacgcaga
17761 ctgttgattc atcacagggt tctgaatatg actatgtcat attcacacaa actactgaaa
17821 cagcacactc ttgtaatgtc aaccgcttca atgtggctat cacaagggca aaaattggca
17881 ttttgtgcat aatgtctgat agagatcttt atgacaaact gcaatttaca agtctagaaa
17941 taccacgtcg caatgtggct acattacaag cagaaaatgt aactggactt tttaaggact
18001 gtagtaagat cattactggt cttcatccta cacaggcacc tacacacctc agcgttgata
18061 taaagttcaa gactgaagga ttatgtgttg acataccagg cataccaaag gacatgacct
18121 accgtagact catctctatg atgggtttca aaatgaatta ccaagtcaat ggttacccta
18181 atatgtttat cacccgcgaa gaagctattc gtcacgttcg tgcgtggatt ggctttgatg
18241 tagagggctg tcatgcaact agagatgctg tgggtactaa cctacctctc cagctaggat
18301 tttctacagg tgttaactta gtagctgtac cgactggtta tgttgacact gaaaataaca
18361 cagaattcac cagagttaat gcaaaacctc caccaggtga ccagtttaaa catcttatac
18421 cactcatgta taaaggcttg ccctggaatg tagtgcgtat taagatagta caaatgctca
18481 gtgatacact gaaaggattg tcagacagag tcgtgttcgt cctttgggcg catggctttg
18541 agcttacatc aatgaagtac tttgtcaaga ttggacctga aagaacgtgt tgtctgtgtg
18601 acaaacgtgc aacttgcttt tctacttcat cagatactta tgcctgctgg aatcattctg
18661 tgggttttga ctatgtctat aacccattta tgattgatgt tcagcagtgg ggctttacgg
18721 gtaaccttca gagtaaccat gaccaacatt gccaggtaca tggaaatgca catgtggcta
18781 gttgtgatgc tatcatgact agatgtttag cagtccatga gtgctttgtt aagcgcgttg
18841 attggtctgt tgaataccct attataggag atgaactgag ggttaattct gcttgcagaa
18901 aagtacaaca catggttgtg aagtctgcat tgcttgctga taagtttcca gttcttcatg
18961 acattggaaa tccaaaggct atcaagtgtg tgcctcaggc tgaagtagaa tggaagttct
19021 acgatgctca gccatgtagt gacaaagctt acaaaataga ggaactcttc tattcttatg
19081 ctacacatca cgataaattc actgatggtg tttgtttgtt ttggaattgt aacgttgatc
19141 gttacccagc caatgcaatt gtgtgtaggt ttgacacaag agtcttgtca aacttgaact
19201 taccaggctg tgatggtggt agtttgtatg tgaataagca tgcattccac actccagctt
19261 tcgataaaag tgcatttact aatttaaagc aattgccttt cttttactat tctgatagtc
19321 cttgtgagtc tcatggcaaa caagtagtgt cggatattga ttatgttcca ctcaaatctg
19381 ctacgtgtat tacacgatgc aatttaggtg gtgctgtttg cagacaccat gcaaatgagt
19441 accgacagta cttggatgca tataatatga tgatttctgc tggatttagc ctatggattt
19501 acaaacaatt tgatacttat aacctgtgga atacatttac caggttacag agtttagaaa
19561 atgtggctta taatgttgtt aataaaggac actttgatgg acacgccggc gaagcacctg
19621 tttccatcat taataatgct gtttacacaa aggtagatgg tattgatgtg gagatctttg
19681 aaaataagac aacacttcct gttaatgttg catttgagct ttgggctaag cgtaacatta
19741 aaccagtgcc agagattaag atactcaata atttgggtgt tgatatcgct gctaatactg
19801 taatctggga ctacaaaaga gaagccccag cacatgtatc tacaataggt gtctgcacaa
19861 tgactgacat tgccaagaaa cctactgaga gtgcttgttc ttcacttact gtcttgtttg
19921 atggtagagt ggaaggacag gtagaccttt ttagaaacgc ccgtaatggt gttttaataa
19981 cagaaggttc agtcaaaggt ctaacacctt caaagggacc agcacaagct agcgtcaatg
20041 gagtcacatt aattggagaa tcagtaaaaa cacagtttaa ctactttaag aaagtagacg

20101 gcattattca acagttgcct gaaacctact ttactcagag cagagactta gaggatttta
20161 agcccagatc acaaatggaa actgactttc tcgagctcgc tatggatgaa ttcatacagc
20221 gatataagct cgagggctat gccttcgaac acatcgttta tggagatttc agtcatggac
20281 aacttggcgg tcttcattta atgataggct tagccaagcg ctcacaagat tcaccactta
20341 aattagagga ttttatccct atggacagca cagtgaaaaa ttacttcata acagatgcgc
20401 aaacaggttc atcaaaatgt gtgtgttctg tgattgatct tttacttgat gactttgtcg
20461 agataataaa gtcacaagat ttgtcagtga tttcaaaagt ggtcaaggtt acaattgact
20521 atgctgaaat ttcattcatg ctttggtgta aggatggaca tgttgaaacc ttctacccaa
20581 aactacaagc aagtcaagcg tggcaaccag gtgttgcgat gcctaacttg tacaagatgc
20641 aaagaatgct tcttgaaaag tgtgaccttc agaattatgg tgaaaatgct gttataccaa
20701 aaggaataat gatgaatgtc gcaaagtata ctcaactgtg tcaatactta aatacactta
20761 ctttagctgt accctacaac atgagagtta ttcactttgg tgctggctct gataaaggag
20821 ttgcaccagg tacagctgtg ctcagacaat ggttgccaac tggcacacta cttgtcgatt
20881 cagatcttaa tgacttcgtc tccgacgcag attctacttt aattggagac tgtgcaacag
20941 tacatacggc taataaatgg gaccttatta ttagcgatat gtatgaccct aggaccaaac
21001 atgtgacaaa agagaatgac tctaaagaag ggtttttcac ttatctgtgt ggatttataa
21061 agcaaaaact agccctgggt ggttctatag ctgtaaagat aacagagcat tcttggaatg
21121 ctgaccttta caagcttatg ggccatttct catggtggac agcttttgtt acaaatgtaa
21181 atgcatcatc atcggaagca tttttaattg ggctaacta tcttggcaag ccgaaggaac
21241 aaattgatgg ctataccatg catgctaact acattttctg gaggaacaca aatcctatcc
21301 agttgtcttc ctattcactc tttgacatga gcaaatttcc tcttaaatta agaggaactg
21361 ctgtaatgtc tcttaaggag aatcaaatca atgatatgat ttattctctt ctggaaaaag
21421 gtaggcttat cattagagaa aacaacagag ttgtggtttc aagtgatatt cttgttaaca

Gene S underscored→

21481 actaaacgaa c**ATG**tttatt ttcttattat ttcttactct cactagtggt agtgaccttg
21541 accggtgcac cacttttgat gatgttcaag ctcctaatta cactcaacat acttcatcta
21601 tgaggggggt ttactatcct gatgaaattt ttagatcaga cactctttat ttaactcagg
21661 atttatttct tccattttat tctaatgtta cagggtttca tactattaat catacgtttg
21721 gcaaccctgt catacctttt aaggatggta tttattttgc tgccacagag aaatcaaatg
21781 ttgtccgtgg ttgggttttt ggttctacca tgaacaacaa gtcacagtcg gtgattatta
21841 ttaacaattc tactaatgtt gttatacgag catgtaactt tgaattgtgt gacaaccctt
21901 tctttgctgt ttctaaaccc atgggtacac agacacatac tatgatattc gataatgcat
21961 ttaattgcac tttcgagtac atatctgatg ccttttcgct tgatgtttca gaaaagtcag
22021 gtaattttaa acacttacga gagtttgtgt ttaaaaataa agatgggttt ctctatgttt
22081 ataagggcta tcaacctata gatgtagttc gtgatctacc ttctggtttt aacactttga
22141 aacctatttt taagttgcct cttggtatta acattacaaa ttttagagcc attcttacag
22201 ccttttcacc tgctcaagac atttggggca cgtcagctgc agcctatttt gttggctatt
22261 taaagccaac tacatttatg ctcaagtatg atgaaaatgg tacaatcaca gatgctgttg
22321 attgttctca aaatccactt gctgaactca aatgctctgt taagagcttt gagattgaca
22381 aaggaattta ccagacctct aatttcaggg ttgttccctc aggagatgtt gtgagattcc
22441 ctaatattac aaacttgtgt ccttttggag aggtttttaa tgctactaaa ttcccttctg
22501 tctatgcatg ggagagaaaa aaaatttcta attgtgttgc tgattactct gtgctctaca
22561 actcaacatt tttttcaacc tttaagtgct atggcgtttc tgccactaag ttgaatgatc
22621 tttgcttctc caatgtctat gcagattctt ttgtagtcaa gggagatgat gtaagacaaa
22681 tagcgccagg acaaactggt gttattgctg attataatta taaattgcca gatgatttca
22741 tgggttgtgt ccttgcttgg aatactagga acattgatgc tacttcaact ggtaattata
22801 attataaata taggtatctt agacatggca agcttaggcc ctttgagaga gacatatcta
22861 atgtgccttt ctcccctgat ggcaaacctt gcaccccacc tgctcttaat tgttattggc
22921 cattaaatga ttatggtttt tacaccacta ctggcattgg ctaccaacct tacagagttg
22981 tagtactttc ttttgaactt ttaaatgcac cggccacggt ttgtggacca aaattatcca
23041 ctgaccttat taagaaccag tgtgtcaatt ttaattttaa tggactcact ggtactggtg
23101 tgttaactcc ttcttcaaag agatttcaac catttcaaca atttggccgt gatgtttctg
23161 atttcactga ttccgttcga gatcctaaaa catctgaaat attagacatt tcaccttgct
23221 cttttggggg tgtaagtgta attacacctg gaacaaatgc ttcatctgaa gttgctgttc
23281 tatatcaaga tgttaactgc actgatgttt ctacagcaat tcatgcagat caactcacac
23341 cagcttggcg catatattct actggaaaca atgtattcca gactcaagca ggctgtctta
23401 taggagctga gcatgtcgac acttcttatg agtgcgacat tcctattgga gctggcattt
23461 gtgctagtta ccatacagtt tctttattac gtagtactag ccaaaaatct attgtggctt
23521 atactatgtc tttaggtgct gatagttcaa ttgcttactc taataacacc attgctatac
23581 ctactaactt ttcaattagc attactacag aagtaatgcc tgtttctatg gctaaaacct
23641 ccgtagattg taatatgtac atctgcggag attctactga atgtgctaat ttgcttctcc
23701 aatatggtag cttttgcaca caactaaatc gtgcactctc aggtattgct gctgaacagg
23761 atcgcaacac acgtgaagtg ttcgctcaag tcaaacaaat gtacaaaacc ccaactttga
23821 aatattttgg tggttttaat ttttcacaaa tattacctga ccctctaaag ccaactaaga
23881 ggtcttttat tgaggacttg ctctttaata aggtgacact cgctgatgct ggcttcatga
23941 agcaatatgg cgaatgccta ggtgatatta atgctagaga tctcatttgt gcgcagaagt

24001 tcaatggact tacagtgttg ccacctctgc tcactgatga tatgattgct gcctacactg
24061 ctgctctagt tagtggtact gccactgctg gatggacatt tggtgctggc gctgctcttc
24121 aaatacctttt tgctatgcaa atggcatata ggttcaatgg cattggagtt acccaaaatg
24181 ttctctatga gaaccaaaaa caaatcgcca accaatttaa caaggcgatt agtcaaattc
24241 aagaatcact tacaacaaca tcaactgcat tgggcaagct gcaagacgtt gttaaccaga
24301 atgctcaagc attaaacaca cttgttaaac aacttagctc taattttggt gcaatttcaa
24361 gtgtgctaaa tgatatcctt tcgcgacttg ataaagtcga ggcggaggta caaattgaca
24421 ggttaattac aggcagactt caaagccttc aaacctatgt aacacaacaa ctaatcaggg
24481 ctgctgaaat cagggcttct gctaatcttg ctgctactaa aatgtctgag tgtgttcttg
24541 gacaatcaaa aagagttgac ttttgtggaa agggctacca ccttatgtcc ttcccacaag
24601 cagccccgca tggtgttgtc ttcctacatg tcacgtatgt gccatcccag gagaggaact
24661 tcaccacagc gccagcaatt tgtcatgaag gcaaagcata cttccctcgt gaaggtgttt
24721 ttgtgtttaa tggcacttct tggtttatta cacagaggaa cttcttttct ccacaaataa
24781 ttactacaga caatacattt gtctcaggaa attgtgatgt cgttattggc atcattaaca
24841 acacagttta tgatcctctg caacctgagc ttgactcatt caaagaagag ctggacaagt
24901 acttcaaaaa tcatacatca ccagatgttg atcttggcga catttcaggc attaacgctt
24961 ctgtcgtcaa cattcaaaaa gaaattgacc gcctcaatga ggtcgctaaa aatttaaatg
25021 aatcactcat tgaccttcaa gaattgggaa aatatgagca atatattaaa tggccttggt
25081 atgtttggct cggcttcatt gctggactaa ttgccatcgt catggttaca atcttgcttt
25141 gttgcatgac tagttgttgc agttgcctca agggtgcatg ctcttgtggt tcttgctgca
25201 agtttgatga ggatgactct gagccagttc tcaagggtgt caaattacat tacaca*TAA*a
25261 cgaacttatg gatttgttta tgagattttt tactcttgga tcaattactg cacagccagt
25321 aaaaattgac aatgcttctc ctgcaagtac tgttcatgct acagcaacga taccgctaca
25381 agcctcactc cctttcggat ggcttgttat tggcgttgca tttcttgctg tttttcagag
25441 cgctaccaaa ataattgcgc tcaataaaag atggcagcta gccctttata agggcttcca
25501 gttcatttgc aatttactgc tgctatttgt taccatctat tcacatcttt tgcttgtcgc
25561 tgcaggtatg gaggcgcaat ttttgtacct ctatgccttg atatattttc tacaatgcat
25621 caacgcatgt agaattatta tgagatgttg gctttgttgg aagtgcaaat ccaagaaccc
25681 attactttat gatgccaact actttgtttg ctggcacaca cataactatg actactgtat
25741 accatataac agtgtcacag atacaattgt cgttactgaa ggtgacggca tttcaacacc
25801 aaaactcaaa gaagactacc aaattggtgg ttattctgag gataggcact caggtgttaa
25861 agactatgtc gttgtacatg gctatttcac cgaagtttac taccagcttg agtctacaca
25921 aattactaca gacactggta ttgaaaatgc tacattcttc atctttaaca agcttgttaa
25981 agacccaccg aatgtgcaaa tacacacaat cgacggctct tcaggagttg ctaatccagc
26041 aatggatcca atttatgatg agccgacgac gactactagc gtgcctttgt aagcacaaga

Gene E underscored→

26101 aagtgagtac gaacttATGt actcattcgt ttcggaagaa acaggtacgt taatagttaa
26161 tagcgtactt ctttttcttg ctttcgtggt attcttgcta gtcacactag ccatccttac
26221 tgcgcttcga ttgtgtgcgt actgctgcaa tattgttaac gtgagtttag taaaaccaac
26281 ggtttacgtc tactcgcgtg ttaaaaatct gaactcttct gaaggagttc ctgatcttct
26341 ggtc*TAA*acg aactaactat tattattatt ctgtttggaa ctttaacatt gcttatc*atg*

<-Gene M underscored→

26401 gcagacaacg gtactattac cgttgaggag cttaaacaac tcctggaaca atggaaccta
26461 gtaataggtt tcctattcct agcctggatt atgttactac aatttgccta ttctaatcgg
26521 aacaggtttt tgtacataat aaagcttgtt ttcctctggc tcttgtggcc agtaacactt
26581 gcttgttttg tgcttgctgc tgtctacaga attaattggg tgactggcgg gattgcgatt
26641 gcaatggctt gtattgtagg cttgatgtgg cttagctact tcgttgcttc cttcaggctg
26701 tttgctcgta cccgctcaat gtggtcattc aacccagaaa caaacattct tctcaatgtg
26761 cctctccggg ggacaattgt gaccagaccg ctcatggaaa gtgaacttgt cattggtgct
26821 gtgatcattc gtggtcactt gcgaatggcc ggacactccc tagggcgctg tgacattaag
26881 gacctgccaa aagagatcac tgtggctaca tcacgaacgc tttcttatta caaattagga
26941 gcgtcgcagc gtgtaggcac tgattcaggt tttgctgcat acaaccgcta ccgtattgga
27001 aactataaat taaatacaga ccacgccggt agcaacgaca atattgcttt gctagtacag
27061 TAAgtgacaa cagatgtttc atcttgttga cttccaggtt acaatagcag agatattgat
27121 tatcattatg aggactttca ggattgctat ttggaatctt gacgttataa taagttcaat
27181 agtgagacaa ttatttaagc ctctaactaa gaagaattat tcggagttag atgatgaaga
27241 acctatggag ttagattatc cataaaacga acatgaaaat tattctcttc ctgacattga
27301 ttgtatttac atcttgcgag ctatatcact atcaggagtg tgttagaggt acgactgtac
27361 tactaaaaga accttgccca tcaggaacat acgagggcaa ttcaccattt caccctcttg
27421 ctgacaataa atttgcacta acttgcacta gcacacactt tgcttttgct tgtgctgacg
27481 gtactcgaca tacctatcag ctgcgtgcaa gatcagtttc accaaaactt ttcatcagac
27541 aagaggaggt tcaacaagag ctctactcgc cactttttct cattgttgct gctctagtat
27601 ttttaatact ttgcttcacc attaagagaa agacagaatg aatgagctca ctttaattga
27661 cttctatttg tgctttttag cctttctgct attccttgtt ttaataatgc ttattatatt
27721 ttggttttca ctcgaaatcc aggatctaga agaaccttgt accaaagtct aaacgaacat
27781 gaaacttctc attgttttga cttgtatttc tctatgcagt tgcatatgca ctgtagtaca

27841 gcgctgtgca tctaataaac ctcatgtgct tgaagatcct tgtaaggtac aacactaggg
27901 gtaatactta tagcactgct tggctttgtg ctctaggaaa ggttttacct tttcatagat
27961 ggcacactat ggttcaaaca tgcacaccta atgttactat caactgtcaa gatccagctg
28021 gtggtgcgct tatagctagg tgttggtacc ttcatgaagg tcaccaaact gctgcattta
<-Gene N underscored->
28081 gagacgtact tgttgtttta aataaacgaa caaattaaa*A TG*tctgataa tggaccccaa
28141 tcaaaccaac gtagtgcccc ccgcattaca tttggtggac ccacagattc aactgacaat
28201 aaccagaatg gaggacgcaa tggggcaagg ccaaaacagc gccgacccca aggtttaccc
28261 aataatactg cgtcttggtt cacagctctc actcagcatg gcaaggagga acttagattc
28321 cctcgaggcc agggcgttcc aatcaacacc aatagtggtc cagatgacca aattggctac
28381 taccgaagag ctacccgacg agttcgtggt ggtgacggca aaatgaaaga gctcagcccc
28441 agatggtact tctattacct aggaactggc ccagaagctt cacttcccta cggcgctaac
28501 aaagaaggca tcgtatgggt tgcaactgag ggagccttga atacacccaa agaccacatt
28561 ggcacccgca atcctaataa caatgctgcc accgtgctac aacttcctca aggaacaaca
28621 ttgccaaaag gcttctacgc agagggaagc agaggcggca gtcaagcctc ttctcgctcc
28681 tcatcacgta gtcgcggtaa ttcaagaaat tcaactcctg gcagcagtag gggaaattct
28741 cctgctcgaa tggctagcgg aggtggtgaa actgccctcg cgctattgct gctagacaga
28801 ttgaaccagc ttgagagcaa agtttctggt aaaggccaac aacaacaagg ccaaactgtc
28861 actaagaaat ctgctgctga ggcatctaaa aagcctcgcc aaaaacgtac tgccacaaaa
28921 cagtacaacg tcactcaagc atttgggaga cgtggtccag aacaaaccca aggaaatttc
28981 ggggaccaag acctaatcag acaaggaact gattacaaac attggccgca aattgcacaa
29041 tttgctccaa gtgcctctgc attctttgga atgtcacgca ttggcatgga agtcacacct
29101 tcgggaacat ggctgactta tcatggagcc attaaattgg atgacaaaga tccacaattc
29161 aaagacaacg tcatactgct gaacaagcac attgacgcat acaaaacatt cccaccaaca
29221 gagcctaaaa aggacaaaaa gaaaaagact gatgaagctc agcctttgcc gcagagacaa
29281 aagaagcagc ccactgtgac tcttcttcct gcggctgaca tggatgattt ctccagacaa
29341 cttcaaaatt ccatgagtgg agcttctgct gattcaactc aggca*TAA*ac actcatgatg
29401 accacacaag gcagatgggc tatgtaaacg ttttcgcaat tccgtttacg atacatagtc
29461 tactcttgtg cagaatgaat tctcgtaact aaacagcaca agtaggttta gttaacttta
29521 atctcacata gcaatcttta atcaatgtgt aacattaggg aggacttgaa agagccacca
29581 cattttcatc gaggccacgc ggagtacgat cgagggtaca gtgaataatg ctagggagag
29641 ctgcctatat ggaagagccc taatgtgtaa aattaatttt agtagtgcta tccccatgtg
29701 attttaatag cttcttagga gaatgacaa

The following subsequences are shown and annotated above by underscoring the coding sequences of interest with the initiation codon ATG in uppercase characters, and the stop codon in uppercase italic characters.

The individual coding sequences and translated amino acid sequences are provided below:

1. The coding sequence for the S (spike) glycoprotein, SEQ ID NO:13, is from nt 21492 to 25259 of SEQ ID NO:12, which comprises 3768 nt that encode 1255 residues + stop codon.

SEQ ID NO:13

ATG ttt att ttc tta tta ttt ctt act ctc act agt ggt agt gac ctt gac cgg tgc
acc act ttt gat gat gtt caa gct cct aat tac act caa cat act tca tct atg agg
ggg gtt tac tat cct gat gaa att ttt aga tca gac act ctt tat tta act cag gat
tta ttt ctt cca ttt tat tct aat gtt aca ggg ttt cat act att aat cat acg ttt
ggc aac cct gtc ata cct ttt aag gat ggt att tat ttt gct gcc aca gag aaa tca
aat gtt gtc cgt ggt tgg gtt ttt ggt tct acc atg aac aac aag tca cag tcg gtg
att att att aac aat tct act aat gtt gtt ata cga gca tgt aac ttt gaa ttg tgt
gac aac cct ttc ttt gct gtt tct aaa ccc atg ggt aca cag aca cat act atg ata
ttc gat aat gca ttt aat tgc act ttc gag tac ata tct gat gcc ttt tcg ctt gat
gtt tca gaa aag tca ggt aat ttt aaa cac tta cga gag ttt gtg ttt aaa aat aaa
gat ggg ttt ctc tat gtt tat aag ggc tat caa cct ata gat gta gtt cgt gat cta
cct tct ggt ttt aac act ttg aaa cct att ttt aag ttg cct ctt ggt att aac att
aca aat ttt aga gcc att ctt aca gcc ttt tca cct gct caa gac att tgg ggc acg
tca gct gca gcc tat ttt gtt ggc tat tta aag cca act aca ttt atg ctc aag tat
gat gaa aat ggt aca atc aca gat gct gtt gat tgt tct caa aat cca ctt gct gaa
ctc aaa tgc tct gtt aag agc ttt gag att gac aaa gga att tac cag acc tct aat

ttc agg gtt gtt ccc tca gga gat gtt gtg aga ttc cct aat att aca aac ttg tgt
cct ttt gga gag gtt ttt aat gct act aaa ttc cct tct gtc tat gca tgg gag aga
aaa aaa att tct aat tgt gtt gct gat tac tct gtg ctc tac aac tca aca ttt ttt
tca acc ttt aag tgc tat ggc gtt tct gcc act aag ttg aat gat ctt tgc ttc tcc
aat gtc tat gca gat tct ttt gta gtc aag gga gat gat gta aga caa ata gcg cca
gga caa act ggt gtt att gct gat tat aat tat aaa ttg cca gat gat ttc atg ggt
tgt gtc ctt gct tgg aat act agg aac att gat gct act tca act ggt aat tat aat
tat aaa tat agg tat ctt aga cat ggc aag ctt agg ccc ttt gag aga gac ata tct
aat gtg cct ttc tcc cct gat ggc aaa cct tgc acc cca cct gct ctt aat tgt tat
tgg cca tta aat gat tat ggt ttt tac acc act act ggc att ggc tac caa cct tac
aga gtt gta gta ctt tct ttt gaa ctt tta aat gca ccg gcc acg gtt tgt gga cca
aaa tta tcc act gac ctt att aag aac cag tgt gtc aat ttt aat ttt aat gga ctc
act ggt act ggt gtg tta act cct tct tca aag aga ttt caa cca ttt caa caa ttt
ggc cgt gat gtt tct gat ttc act gat tcc gtt cga gat cct aaa aca tct gaa ata
tta gac att tca cct tgc tct ttt ggg ggt gta agt gta att aca cct gga aca aat
gct tca tct gaa gtt gct gtt cta tat caa gat gtt aac tgc act gat gtt tct aca
gca att cat gca gat caa ctc aca cca gct tgg cgc ata tat tct act gga aac aat
gta ttc cag act caa gca ggc tgt ctt ata gga gct gag cat gtc gac act tct tat
gag tgc gac att cct att gga gct ggc att tgt gct agt tac cat aca gtt tct tta
tta cgt agt act agc caa aaa tct att gtg gct tat act atg tct tta ggt gct gat
agt tca att gct tac tct aat aac acc att gct ata cct act aac ttt tca att agc
att act aca gaa gta atg cct gtt tct atg gct aaa acc tcc gta gat tgt aat atg
tac atc tgc gga gat tct act gaa tgt gct aat ttg ctt ctc caa tat ggt agc ttt
tgc aca caa cta aat cgt gca ctc tca ggt att gct gct gaa cag gat cgc aac aca
cgt gaa gtg ttc gct caa gtc aaa caa atg tac aaa acc cca act ttg aaa tat ttt
ggt ggt ttt aat ttt tca caa ata tta cct gac cct cta aag cca act aag agg tct
ttt att gag gac ttg ctc ttt aat aag gtg aca ctc gct gat gct ggc ttc atg aag
caa tat ggc gaa tgc cta ggt gat att aat gct aga gat ctc att tgt gcg cag aag
ttc aat gga ctt aca gtg ttg cca cct ctg ctc act gat gat atg att gct gcc tac
act gct gct cta gtt agt ggt act gcc act gct gga tgg aca ttt ggt gct ggc gct
gct ctt caa ata cct ttt gct atg caa atg gca tat agg ttc aat ggc att gga gtt
acc caa aat gtt ctc tat gag aac caa aaa caa atc gcc aac caa ttt aac aag gcg
att agt caa att caa gaa tca ctt aca aca aca tca act gca ttg ggc aag ctg caa
gac gtt gtt aac cag aat gct caa gca tta aac aca ctt gtt aaa caa ctt agc tct
aat ttt ggt gca att tca agt gtg cta aat gat atc ctt tcg cga ctt gat aaa gtc
gag gcg gag gta caa att gac agg tta att aca ggc aga ctt caa agc ctt caa acc
tat gta aca caa caa cta atc agg gct gct gaa atc agg gct tct gct aat ctt gct
gct act aaa atg tct gag tgt gtt ctt gga caa tca aaa aga gtt gac ttt tgt gga
aag ggc tac cac ctt atg tcc ttc cca caa gca gcc ccg cat ggt gtt gtc ttc cta
cat gtc acg tat gtg cca tcc cag gag agg aac ttc acc aca gcg cca gca att tgt
cat gaa ggc aaa gca tac ttc cct cgt gaa ggt gtt ttt gtg ttt aat ggc act tct
tgg ttt att aca cag agg aac ttc ttt tct cca caa ata att act aca gac aat aca
ttt gtc tca gga aat tgt gat gtc gtt att ggc atc att aac aac aca gtt tat gat
cct ctg caa cct gag ctt gac tca ttc aaa gaa gag ctg gac aag tac ttc aaa aat
cat aca tca cca gat gtt gat ctt ggc gac att tca ggc att aac gct tct gtc gtc
aac att caa aaa gaa att gac cgc ctc aat gag gtc gct aaa aat tta aat gaa tca
ctc att gac ctt caa gaa ttg gga aaa tat gag caa tat att aaa tgg cct tgg tat
gtt tgg ctc ggc ttc att gct gga cta att gcc atc gtc atg gtt aca atc ttg ctt
tgt tgc atg act agt tgt tgc agt tgc ctc aag ggt gca tgc tct tgt ggt tct tgc
tgc aag ttt gat gag

The encoded amino acid sequence of the S polypeptide (SEQ ID NO:14) is:

MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL 60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS 120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK 180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP 240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY 300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF 360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV 420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND 480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP 540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD 600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY 660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC 720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG 780

```
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL   840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE   900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN   960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK  1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN  1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN  1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL  1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT        1255
```

Sequences of domains of the S polypeptide (see Figure 6) are set forth below:

Domain S1: – amino acids 1-680 of SEQ ID NO:14 which is shown below as

SEQ ID NO:15:

```
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL    60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS   120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK   180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP   240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY   300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF   360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV   420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND   480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP   540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD   600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY   660
HTVSLLRSTS QKSIVAYTMS                                               680
```

Domain S2 - aa 680-1225 of SEQ ID NO:14 which is shown below as SEQ ID NO:16 (residues 1-575):

```
LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC NMYICGDSTE CANLLLQYGS    60
FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG GFNFSQILPD PLKPTKRSFI   120
EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL TVLPPLLTDD MIAAYTAALV   180
SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE NQKQIANQFN KAISQIQESL   240
TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN DILSRLDKVE AEVQIDRLIT   300
GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK RVDFCGKGYH LMSFPQAAPH   360
GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN GTSWFITQRN FFSPQIITTD   420
NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN HTSPDVDLGD ISGINASVVN   480
IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL GFIAGLIAIV MVTILLCCMT   540
SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT                              575
```

Polypeptide Si overlaps domains S1 and S2 and corresponds to residues 417-816 or SEQ ID NO:14. This polypeptide is shown below as SEQ ID:17 (aa 1-400):

```
MGCVLAWNTR NIDATSTGNY NYKYRYLRHG KLRPFERDIS NVPFSPDGKP CTPPALNCYW    60
PLNDYGFYTT TGIGYQPYRV VVLSFELLNA PATVCGPKLS TDLIKNQCVN FNFNGLTGTG   120
VLTPSSKRFQ PFQQFGRDVS DFTDSVRDPK TSEILDISPC SFGGVSVITP GTNASSEVAV   180
LYQDVNCTDV STAIHADQLT PAWRIYSTGN NVFQTQAGCL IGAEHVDTSY ECDIPIGAGI   240
CASYHTVSLL RSTSQKSIVA YTMSLGADSS IAYSNNTIAI PTNFSISITT EVMPVSMAKT   300
SVDCNMYICG DSTECANLLL QYGSFCTQLN RALSGIAAEQ DRNTREVFAQ VKQMYKTPTL   360
KYFGGFNFSQ ILPDPLKPTK RSFIEDLLFN KVTLADAGFM                         400
```

The present invention includes homologous sequences to the S polypeptide domains from any other strain of SARS-CoV.

2. The coding sequence for the E (envelope, or "small envelope") protein (SEQ ID NO:18) is from nt 26117 to 26347 of SEQ ID NO:12, which comprises 231 nt that encode 76 aa's + stop codon.

SEQ ID NO:18

```
ATG tac tca ttc gtt tcg gaa gaa aca ggt acg tta ata gtt aat agc gta
ctt ctt ttt ctt gct ttc gtg gta ttc ttg cta gtc aca cta gcc atc ctt
act gcg ctt cga ttg tgt gcg tac tgc tgc aat att gtt aac gtg agt tta
gta aaa cca acg gtt tac gtc tac tcg cgt gtt aaa aat ctg aac tct tct
gaa gga gtt cct gat ctt ctg gtc TAA
```

The encoded amino acid sequence of the E polypeptide (SEQ ID NO:19) is:

```
MYSFVSEETG TLIVNSVLLF LAFVVFLLVT LAILTALRLC AYCCNIVNVS LVKPTVYVYS   60
RVKNLNSSEG VPDLLV                                                    76
```

3. The coding sequence for the M (membrane protein (SEQ ID NO:20) is from nt 26398 to 27063 of SEQ ID NO:12, which comprises 666 nt encoding 221 aa + stop codon.

SEQ ID NO:20

```
ATG gca gac aac ggt act att acc gtt gag gag ctt aaa caa ctc ctg gaa
caa tgg aac cta gta ata ggt ttc cta ttc cta gcc tgg att atg tta cta
caa ttt gcc tat tct aat cgg aac agg ttt ttg tac ata ata aag ctt gtt
ttc ctc tgg ctc ttg tgg cca gta aca ctt gct tgt ttt gtg ctt gct gct
gtc tac aga att aat tgg gtg act ggc ggg att gcg att gca atg gct tgt
att gta ggc ttg atg tgg ctt agc tac ttc gtt gct tcc ttc agg ctg ttt
gct cgt acc cgc tca atg tgg tca ttc aac cca gaa aca aac att ctt ctc
aat gtg cct ctc cgg ggg aca att gtg acc aga ccg ctc atg gaa agt gaa
ctt gtc att ggt gct gtg atc att cgt ggt cac ttg cga atg gcc gga cac
tcc cta ggg cgc tgt gac att aag gac ctg cca aaa gag atc act gtg gct
aca tca cga acg ctt tct tat tac aaa tta gga gcg tcg cag cgt gta ggc
act gat tca ggt ttt gct gca tac aac cgc tac cgt att gga aac tat aaa
tta aat aca gac cac gcc ggt agc aac gac aat att gct ttg cta gta cag
TAA
```

The encoded amino acid sequence of the M polypeptide (SEQ ID NO:21) is:

```
MADNGTITVE ELKQLLEQWN LVIGFLFLAW IMLLQFAYSN RNRFLYIIKL VFLWLLWPVT   60
LACFVLAAVY RINWVTGGIA IAMACIVGLM WLSYFVASFR LFARTRSMWS FNPETNILLN  120
VPLRGTIVTR PLMESELVIG AVIIRGHLRM AGHSLGRCDI KDLPKEITVA TSRTLSYYKL  180
GASQRVGTDS GFAAYNRYRI GNYKLNTDHA GSNDNIALLV Q                      221
```

4. The coding sequence for the N (nucleocapsid protein (SEQ ID NO:22) is from nt 28120 to 29388 of SEQ ID NO:12, which comprises 1269 nt encoding 422 aa + stop codon.

SEQ ID NO:22

```
ATG tct gat aat gga ccc caa tca aac caa cgt agt gcc ccc cgc att aca ttt ggt
gga ccc aca gat tca act gac aat aac cag aat gga gga cgc aat ggg gca agg cca
aaa cag cgc cga ccc caa ggt tta ccc aat aat act gcg tct tgg ttc aca gct ctc
act cag cat ggc aag gag gaa ctt aga ttc cct cga ggc cag ggc gtt cca atc aac
acc aat agt ggt cca gat gac caa att ggc tac tac cga aga gct acc cga cga gtt
cgt ggt ggt gac ggc aaa atg aaa gag ctc agc ccc aga tgg tac ttc tat tac cta
gga act ggc cca gaa gct tca ctt ccc tac ggc gct aac aaa gaa ggc atc gta tgg
gtt gca act gag gga gcc ttg aat aca ccc aaa gac cac att ggc acc cgc aat cct
aat aac aat gct gcc acc gtg cta caa ctt cct caa gga aca aca ttg cca aaa ggc
```

```
ttc tac gca gag gga agc aga ggc ggc agt caa gcc tct tct cgc tcc tca tca cgt
agt cgc ggt aat tca aga aat tca act cct ggc agc agt agg gga aat tct cct gct
cga atg gct agc gga ggt ggt gaa act gcc ctc gcg cta ttg ctg cta gac aga ttg
aac cag ctt gag agc aaa gtt tct ggt aaa ggc caa caa caa caa ggc caa act gtc
act aag aaa tct gct gct gag gca tct aaa aag cct cgc caa aaa cgt act gcc aca
aaa cag tac aac gtc act caa gca ttt ggg aga cgt ggt cca gaa caa acc caa gga
aat ttc ggg gac caa gac cta atc aga caa gga act gat tac aaa cat tgg ccg caa
att gca caa ttt gct cca agt gcc tct gca ttc ttt gga atg tca cgc att ggc atg
gaa gtc aca cct tcg gga aca tgg ctg act tat cat gga gcc att aaa ttg gat gac
aaa gat cca caa ttc aaa gac aac gtc ata ctg ctg aac aag cac att gac gca tac
aaa aca ttc cca cca aca gag cct aaa aag gac aaa aag aaa aag act gat gaa gct
cag cct ttg ccg cag aga caa aag aag cag ccc act gtg act ctt ctt cct gcg gct
gac atg gat gat ttc tcc aga caa ctt caa aat tcc atg agt gga gct tct gct gat
tca act cag gca TAA
```

The encoded amino acid sequence of the N polypeptide (SEQ ID NO:23) is:

```
MSDNGPQSNQ RSAPRITFGG PTDSTDNNQN GGRNGARPKQ RRPQGLPNNT ASWFTALTQH  60
GKEELRFPRG QGVPINTNSG PDDQIGYYRR ATRRVRGGDG KMKELSPRWY FYYLGTGPEA 120
SLPYGANKEG IVWVATEGAL NTPKDHIGTR NPNNNAATVL QLPQGTTLPK GFYAEGSRGG 180
SQASSRSSSR SRGNSRNSTP GSSRGNSPAR MASGGGETAL ALLLLDRLNQ LESKVSGKGQ 240
QQQGQTVTKK SAAEASKKPR QKRTATKQYN VTQAFGRRGP EQTQGNFGDQ DLIRQGTDYK 300
HWPQIAQFAP SASAFFGMSR IGMEVTPSGT WLTYHGAIKL DDKDPQFKDN VILLNKHIDA 360
YKTFPPTEPK KDKKKKTDEA QPLPQRQKKQ PTVTLLPAAD MDDFSRQLQN SMSGASADST 420
QA                                                                422
```

pcDNA3-CRT/N (SEQ ID NO:24)

Vector sequence (UPPERCASE)

CRT: *lower case/italic*

N protein: **lower case/bold/underscored**

```
      |   10     |   20     |   30     |   40     |   50     |   60     |   70     |   80     |
   1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGTAT 80
  81 CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA 160
 161 CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT 240
 241 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA 320
 321 CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT 400
 401 AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT 480
 481 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA 560
 561 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA 640
 641 TGGGCGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC 720
 721 AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG 800
 801 GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTACTG GCTTATCGAA ATTAATACGA CTCACTATAG 880
 881 GGAGACCCAA GCTGGCTAGC GTTTAAACGG GCCCTCTAGA atgctgctcc ctgtgccgct gctgctcggc ctgctcggcc 960
 961 tggccgccgc cgagcccgtc gtctacttca aggagcagtt tctggacgga gatgggtgga ccgagcgctg gatcgaatcc 1040
1041 aaacacaagt ccgatttggg caaattcgtc ctcagttcgg gcaagttcta cggcgatcag gagaaagata aagggctgca 1120
1121 gaccagccag gacgcccgct tctacgccct gtcggcccga ttcgagccgt tcagcaacaa gggccagcca ctggtggtgc 1200
1201 agttcaccgt gaaacacgag cagaacattg actgcggggg cggctacgtg aagctgtttc cggccggcct ggaccagaag 1280
1281 gacatgcacg gggactctga gtacaacatc atgtttggtc ctgacatctg tggccccggc accaagaagg ttcacgtcat 1360
1361 cttcaactac aagggcaaga acgtgctgat caacaaggac atccgttgca aggacgacga gttcacacac ctgtacacgc 1440
1441 tgatcgtgcg gccggacaac acgtatgagg tgaagattga caacagccag gtggagtcgg gctccctgga ggatgactgg 1520
1521 gacttcctac cccccaagaa gataaaggac ccagatgcct cgaagcctga agactgggac gagcggcca agatcgacga 1600
1601 ccccacggac tccaagcccg aggactggga caagcccgag cacatccccg acccggacgc gaagaagccc gaagactggg 1680
1681 acgaagaaat ggacggagag tgggagccgc cggtgattca gaaccccgag tacaagggtg agtggaagcc gcggcagatc 1760
1761 gacaaccccg attacaaagg cacctggatc caccccgaaa tcgacaaccc cgagtactcg cccgacgcta acatctatgc 1840
1841 ctacgacagc tttgccgtgc tgggcttgga cctctggcag gtcaagtcgg gcaccatctt cgacaacttc ctcatcacca 1920
1921 acgatgaggc gtacgcagag gagtttggca acgagacgtg ggcgtcacc aagacggccg agaagcagat gaaagacaag 2000
2001 caggacgagg agcagcggct gaaggaggag gaggaggaga agaagcggaa ggaggaggag gaggccgagg aggacgagga 2080
2081 ggacaaggac gacaaggagg acgaggatga ggacgaggag gacaaggacg aggaggagga ggaggcggcc gccggccagg 2160
2161 ccaaggacga gctgtagGAA TTCatgtctg ataatggacc ccaatcaaac caacgtagtg ccccccgcat tacatttggt 2240
2241 ggacccacag attcaactga caataaccag aatggaggac gcaatggggc aaggccaaaa cagcgccgac cccaaggttt 2320
2321 acccaataat actgcgtctt ggttcacagc tctcactcag catggcaagg aggaacttag attccctcga ggccagggcg 2400
2401 ttccaatcaa caccaatagt ggtccagatg accaaattgg ctactaccga agagctaccc gacgagttcg tggtggtgac 2480
2481 ggcaaaatga aagagctcag ccccagatgg tacttctatt acctaggaac tggcccagaa gcttcacttc cctacggcgc 2560
2561 taacaaagaa ggcatcgtat gggttgcaac tgagggagcc ttgaatacac ccaaagacca cattggcacc cgcaatccta 2640
2641 ataacaatgc tgccaccgtg ctacaacttc ctcaaggaac aacattgcca aaaggcttct acgcagaggg aagcagaggc 2720
2721 ggcagtcaag cctcttctcg ctcctcatca cgtagtcgcg gtaattcaag aaattcaact cctggcagca gtaggggaaa 2800
```

```
2801 ttctcctgct cgaatggcta gcggaggtgg tgaaactgcc ctcgcgctat tgctgctaga cagattgaac cagcttgaga 2880
2881 gcaaagtttc tggtaaaggc caacaacaac aaggccaaac tgtcactaag aaatctgctg ctgaggcatc taaaaagcct 2960
2961 cgccaaaaac gtactgccac aaaacagtac aacgtcactc aagcatttgg gagacgtggt ccagaacaaa cccaaggaaa 3040
3041 tttcggggac caagacctaa tcagacaagg aactgattac aaacattggc cgcaaattgc acaatttgct ccaagtgcct 3120
3121 ctgcattctt tggaatgtca cgcattggca tggaagtcac accttcggga acatggctga cttatcatgg agccattaaa 3200
3201 ttggatgaca aagatccaca attcaaagac aacgtcatac tgctgaacaa gcacattgac gcatacaaaa cattcccacc 3280
3281 aacagagcct aaaaaggaca aaaagaaaaa gactgatgaa gctcagcctt tgccgcagag acaaaagaag cagcccactg 3360
3361 tgactcttct tcctgcggct gacatggatg atttctccag acaacttcaa aattccatga gtggagcttc tgctgattca 3440
3441 actcaggcaG GTACCAAGCT TGGGCCCGAA CAAAAACTCA TCTCAGAAGA GGATCTGAAT AGCGCCGTCG ACCATCATCA 3520
3521 TCATCATCAT TGAGTTTAAA CGGTCTCCAG CTTAAGTTTA AACCGCTGAT CAGCCTCGAC TGTGCCTTCT AGTTGCCAGC 3600
3601 CATCTGTTGT TTGCCCCTCC CCCGTGCCTT CCTTGACCCT GGAAGGTGCC ACTCCCACTG TCCTTTCCTA ATAAAATGAG 3680
3681 GAAATTGCAT CGCATTGTCT GAGTAGGTGT CATTCTATTC TGGGGGGTGG GGTGGGGCAG GACAGCAAGG GGGAGGATTG 3760
3761 GGAAGACAAT AGCAGGCATG CTGGGGATGC GGTGGGCTCT ATGGCTTCTG AGGCGGAAAG AACCAGCTGG GGCTCTAGGG 3840
3841 GGTATCCCCA CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC 3920
3921 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCCCCGTC AAGCTCTAAA 4000
4001 TCGGGGGCTC CCTTTAGGGT TCCGATTTAG TGCTTTACGG CACCTCGACC CCAAAAAACT TGATTAGGGT GATGGTTCAC 4080
4081 GTAGTGGGCC ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG TCCACGTTCT TTAATAGTGG ACTCTTGTTC 4160
4161 CAAACTGGAA CAACACTCAA CCCTATCTCG GTCTATTCTT TTGATTTATA AGGGATTTTG CCGATTTCGG CCTATTGGTT 4240
4241 AAAAAATGAG CTGATTTAAC AAAAATTTAA CGCGAATTAA TTCTGTGGAA TGTGTGTCAG TTAGGGTGTG GAAAGTCCCC 4320
4321 AGGCTCCCCA GCAGGCAGAA GTATGCAAAG CATGCATCTC AATTAGTCAG CAACCAGGTG TGGAAAGTCC CCAGGCTCCC 4400
4401 CAGCAGGCAG AAGTATGCAA AGCATGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC 4480
4481 CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT GCAGAGGCCG AGGCCGCCTC 4560
4561 TGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT GGAGGCCTAG GCTTTTGCAA AAAGCTCCCG GGAGCTTGTA 4640
4641 TATCCATTTT CGGATCTGAT CAAGAGACAG GATGAGGATC GTTTCGCATG ATTGAACAAG ATGGATTGCA CGCAGGTTCT 4720
4721 CCGGCCGCTT GGGTGGAGAG GCTATTCGGC TATGACTGGG CACAACAGAC AATCGGCTGC TCTGATGCCG CCGTGTTCCG 4800
4801 GCTGTCAGCG CAGGGGCGCC CGGTTCTTTT TGTCAAGACC GACCTGTCCG GTGCCCTGAA TGAACTGCAG GACGAGGCAG 4880
4881 CGCGGCTATC GTGGCTGGCC ACGACGGGCG TTCCTTGCGC AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG 4960
4961 CTGCTATTGG GCGAAGTGCC GGGGCAGGAT CTCCTGTCAT CTCACCTTGC TCCTGCCGAG AAAGTATCCA TCATGGCTGA 5040
5041 TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC ACCAAGCGAA ACATCGCATC GAGCGAGCAC 5120
5121 GTACTCGGAT GGAAGCCGGT CTTGTCGATC AGGATGATCT GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTC 5200
5201 GCCAGGCTCA AGGCGCGCAT GCCCGACGGC GAGGATCTCG TCGTGACCCA TGGCGATGCC TGCTTGCCGA ATATCATGGT 5280
5281 GGAAAATGGC CGCTTTTCTG GATTCATCGA CTGTGGCCGG CTGGGTGTGG CGGACCGCTA TCAGGACATA GCGTTGGCTA 5360
5361 CCCGTGATAT TGCTGAAGAG CTTGGCGGCG AATGGGCTGA CCGCTTCCTC GTGCTTTACG GTATCGCCGC TCCCGATTCG 5440
5441 CAGCGCATCG CCTTCTATCG CCTTCTTGAC GAGTTCTTCT GAGCGGGACT CTGGGGTTCG AAATGACCGA CCAAGCGACG 5520
5521 CCCAACCTGC CATCACGAGA TTTCGATTCC ACCGCCGCCT TCTATGAAAG GTTGGGCTTC GGAATCGTTT TCCGGGACGC 5600
5601 CGGCTGGATG ATCCTCCAGC GCGGGGATCT CATGCTGGAG TTCTTCGCCC ACCCCAACTT GTTTATTGCA GCTTATAATG 5680
5681 GTTACAAATA AAGCAATAGC ATCACAAATT TCACAAATAA AGCATTTTTT TCACTGCATT CTAGTTGTGG TTTGTCCAAA 5760
5761 CTCATCAATG TATCTTATCA TGTCTGTATA CCGTCGACCT CTAGCTAGAG CTTGGCGTAA TCATGGTCAT AGCTGTTTCC 5840
5841 TGTGTGAAAT TGTTATCCGC TCACAATTCC ACACAACATA CGAGCCGGAA GCATAAAGTG TAAAGCCTGG GGTGCCTAAT 5920
5921 GAGTGAGCTA ACTCACATTA ATTGCGTTGC GCTCACTGCC CGCTTTCCAG TCGGGAAACC TGTCGTGCCA GCTGCATTAA 6000
6001 TGAATCGGCC AACGCGCGGG GAGAGGCGGT TTGCGTATTG GGCGCTCTTC CGCTTCCTCG CTCACTGACT CGCTGCGCTC 6080
6081 GGTCGTTCGG CTGCGGCGAG CGGTATCAGC TCACTCAAAG GCGGTAATAC GGTTATCCAC AGAATCAGGG GATAACGCAG 6160
6161 GAAAGAACAT GTGAGCAAAA GGCCAGCAAA AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC TGGCGTTTTT CCATAGGCTC 6240
6241 CGCCCCCCTG ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA GATACCAGGC 6320
```

```
6321 GTTTCCCCCT GGAAGCTCCC TCGTGCGCTC TCCTGTTCCG ACCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT 6400
6401 CGGGAAGCGT GGCGCTTTCT CATAGCTCAC GCTGTAGGTA TCTCAGTTCG GTGTAGGTCG TTCGCTCCAA GCTGGGCTGT 6480
6481 GTGCACGAAC CCCCCGTTCA GCCCGACCGC TGCGCCTTAT CCGGTAACTA TCGTCTTGAG TCCAACCCGG TAAGACACGA 6560
6561 CTTATCGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC AGAGCGAGGT ATGTAGGCGG TGCTACAGAG TTCTTGAAGT 6640
6641 GGTGGCCTAA CTACGGCTAC ACTAGAAGAA CAGTATTTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA 6720
6721 GTTGGTAGCT CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTTT TTTTGTTTGC AAGCAGCAGA TTACGCGCAG 6800
6801 AAAAAAAGGA TCTCAAGAAG ATCCTTTGAT CTTTTCTACG GGGTCTGACG CTCAGTGGAA CGAAAACTCA CGTTAAGGGA 6880
6881 TTTTGGTCAT GAGATTATCA AAAAGGATCT TCACCTAGAT CCTTTTAAAT TAAAAATGAA GTTTTAAATC AATCTAAAGT 6960
6961 ATATATGAGT AAACTTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC ACCTATCTCA GCGATCTGTC TATTTCGTTC 7040
7041 ATCCATAGTT GCCTGACTCC CCGTCGTGTA GATAACTACG ATACGGGAGG GCTTACCATC TGGCCCCAGT GCTGCAATGA 7120
7121 TACCGCGAGA CCCACGCTCA CCGGCTCCAG ATTTATCAGC AATAAACCAG CCAGCCGGAA GGGCCGAGCG CAGAAGTGGT 7200
7201 CCTGCAACTT TATCCGCCTC CATCCAGTCT ATTAATTGTT GCCGGGAAGC TAGAGTAAGT AGTTCGCCAG TTAATAGTTT 7280
7281 GCGCAACGTT GTTGCCATTG CTACAGGCAT CGTGGTGTCA CGCTCGTCGT TTGGTATGGC TTCATTCAGC TCCGGTTCCC 7360
7361 AACGATCAAG GCGAGTTACA TGATCCCCCA TGTTGTGCAA AAAAGCGGTT AGCTCCTTCG GTCCTCCGAT CGTTGTCAGA 7440
7441 AGTAAGTTGG CCGCAGTGTT ATCACTCATG GTTATGGCAG CACTGCATAA TTCTCTTACT GTCATGCCAT CCGTAAGATG 7520
7521 CTTTTCTGTG ACTGGTGAGT ACTCAACCAA GTCATTCTGA GAATAGTGTA TGCGGCGACC GAGTTGCTCT TGCCCGGCGT 7600
7601 CAATACGGGA TAATACCGCG CCACATAGCA GAACTTTAAA AGTGCTCATC ATTGGAAAAC GTTCTTCGGG GCGAAAACTC 7680
7681 TCAAGGATCT TACCGCTGTT GAGATCCAGT TCGATGTAAC CCACTCGTGC ACCCAACTGA TCTTCAGCAT CTTTTACTTT 7760
7761 CACCAGCGTT TCTGGGTGAG CAAAAACAGG AAGGCAAAAT GCCGCAAAAA AGGGAATAAG GGCGACACGG AAATGTTGAA 7840
7841 TACTCATACT CTTCCTTTTT CAATATTATT GAAGCATTTA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT 7920
7921 ATTTAGAAAA ATAAACAAAT AGGGGTTCCG CGCACATTTC CCCGAAAAGT GCCACCTGAC GTC                   7983
     |    10    |    20    |    30    |    40    |    50    |    60    |    70    |    80    |
```

pcDNA3-S (Spike) (SEQ ID NO:25)

Vector sequence, pcDNA3.1 (+) in UPPERCASE

Spike(S) protein sequence (lower case/ bold/<u>underscored</u>)

```
     |    10    |    20    |    30    |    40    |    50    |    60    |    70    |    80    |
   1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGTAT 80
  81 CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA 160
 161 CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT 240
 241 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA 320
 321 CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT 400
 401 AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT 480
 481 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA 560
 561 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA 640
 641 TGGGCGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC 720
 721 AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG 800
 801 GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTACTG GCTTATCGAA ATTAATACGA CTCACTATAG 880
 881 GGAGACCCAA GCTGGCTAGC GTTTAAACTT AAGCTTGGTA CCGAGCTCGG ATCCatgttt attttcttat tatttcttac 960
```

961	tctcactagt	ggtagtgacc	ttgaccggtg	caccactttt	gatgatgttc	aagctcctaa	ttacactcaa	catacttcat	1040
1041	ctatgagggg	ggtttactat	cctgatgaaa	tttttagatc	agacactctt	tatttaactc	aggatttatt	tcttccattt	1120
1121	tattctaatg	ttacagggtt	tcatactatt	aatcatacgt	ttggcaaccc	tgtcatacct	tttaaggatg	gtatttattt	1200
1201	tgctgccaca	gagaaatcaa	atgttgtccg	tggttgggtt	tttggttcta	ccatgaacaa	caagtcacag	tcggtgatta	1280
1281	ttattaacaa	ttctactaat	gttgttatac	gagcatgtaa	ctttgaattg	tgtgacaacc	ctttctttgc	tgtttctaaa	1360
1361	cccatgggta	cacagacaca	tactatgata	ttcgataatg	catttaattg	cactttcgag	tacatatctg	atgccttttc	1440
1441	gcttgatgtt	tcagaaaagt	caggtaattt	taaacactta	cgagagtttg	tgtttaaaaa	taaagatggg	tttctctatg	1520
1521	tttataaggg	ctatcaacct	atagatgtag	ttcgtgatct	accttctggt	tttaacactt	tgaaacctat	ttttaagttg	1600
1601	cctcttggta	ttaacattac	aaattttaga	gccattctta	cagccttttc	acctgctcaa	gacatttggg	gcacgtcagc	1680
1681	tgcagcctat	tttgttggct	atttaaagcc	aactacattt	atgctcaagt	atgatgaaaa	tggtacaatc	acagatgctg	1760
1761	ttgattgttc	tcaaaatcca	cttgctgaac	tcaaatgctc	tgttaagagc	tttgagattg	acaaaggaat	ttaccagacc	1840
1841	tctaatttca	gggttgttcc	ctcaggagat	gttgtgagat	tccctaatat	tacaaacttg	tgtccttttg	gagaggtttt	1920
1921	taatgctact	aaattccctt	ctgtctatgc	atgggagaga	aaaaaaattt	ctaattgtgt	tgctgattac	tctgtgctct	2000
2001	acaactcaac	atttttttca	acctttaagt	gctatggcgt	ttctgccact	aagttgaatg	atctttgctt	ctccaatgtc	2080
2081	tatgcagatt	cttttgtagt	caagggagat	gatgtaagac	aaatagcgcc	aggacaaact	ggtgttattg	ctgattataa	2160
2161	ttataaattg	ccagatgatt	tcatgggttg	tgtccttgct	tggaatacta	ggaacattga	tgctacttca	actggtaatt	2240
2241	ataattataa	atataggtat	cttagacatg	gcaagcttag	gccctttgag	agagacatat	ctaatgtgcc	tttctcccct	2320
2321	gatggcaaac	cttgcacccc	acctgctctt	aattgttatt	ggccattaaa	tgattatggt	ttttacacca	ctactggcat	2400
2401	tggctaccaa	ccttacagag	ttgtagtact	ttcttttgaa	cttttaaatg	caccggccac	ggtttgtgga	ccaaaattat	2480
2481	ccactgacct	tattaagaac	cagtgtgtca	attttaattt	taatggactc	actggtactg	gtgtgttaac	tccttcttca	2560
2561	aagagatttc	aaccatttca	acaatttggc	cgtgatgttt	ctgatttcac	tgattccgtt	cgagatccta	aaacatctga	2640
2641	aatattagac	atttcacctt	gctcttttgg	gggtgtaagt	gtaattacac	ctggaacaaa	tgcttcatct	gaagttgctg	2720
2721	ttctatatca	agatgttaac	tgcactgatg	tttctacagc	aattcatgca	gatcaactca	caccagcttg	gcgcatatat	2800
2801	tctactggaa	acaatgtatt	ccagactcaa	gcaggctgtc	ttataggagc	tgagcatgtc	gacacttctt	atgagtgcga	2880
2881	cattcctatt	ggagctggca	tttgtgctag	ttaccataca	gtttctttat	tacgtagtac	tagccaaaaa	tctattgtgg	2960
2961	cttatactat	gtctttaggt	gctgatagtt	caattgctta	ctctaataac	accattgcta	tacctactaa	cttttcaatt	3040
3041	agcattacta	cagaagtaat	gcctgtttct	atggctaaaa	cctccgtaga	ttgtaatatg	tacatctgcg	gagattctac	3120
3121	tgaatgtgct	aatttgcttc	tccaatatgg	tagcttttgc	acacaactaa	atcgtgcact	ctcaggtatt	gctgctgaac	3200
3201	aggatcgcaa	cacacgtgaa	gtgttcgctc	aagtcaaaca	aatgtacaaa	accccaactt	tgaaatattt	tggtggtttt	3280
3281	aatttttcac	aaatattacc	tgaccctcta	aagccaacta	agaggtcttt	tattgaggac	ttgctcttta	ataaggtgac	3360
3361	actcgctgat	gctggcttca	tgaagcaata	tggcgaatgc	ctaggtgata	ttaatgctag	agatctcatt	tgtgcgcaga	3440
3441	agttcaatgg	acttacagtg	ttgccacctc	tgctcactga	tgatatgatt	gctgcctaca	ctgctgctct	agttagtggt	3520
3521	actgccactg	ctggatggac	atttggtgct	ggcgctgctc	ttcaaatacc	ttttgctatg	caaatggcat	ataggttcaa	3600
3601	tggcattgga	gttacccaaa	atgttctcta	tgagaaccaa	aaacaaatcg	ccaaccaatt	taacaaggcg	attagtcaaa	3680
3681	ttcaagaatc	acttacaaca	acatcaactg	cattgggcaa	gctgcaagac	gttgttaacc	agaatgctca	agcattaaac	3760
3761	acacttgtta	aacaacttag	ctctaatttt	ggtgcaattt	caagtgtgct	aaatgatatc	ctttcgcgac	ttgataaagt	3840
3841	cgaggcggag	gtacaaattg	acaggttaat	tacaggcaga	cttcaaagcc	ttcaaaccta	tgtaacacaa	caactaatca	3920
3921	gggctgctga	aatcagggct	tctgctaatc	ttgctgctac	taaaatgtct	gagtgtgttc	ttggacaatc	aaaaagagtt	4000
4001	gacttttgtg	gaaagggcta	ccaccttatg	tccttcccac	aagcagcccc	gcatggtgtt	gtcttcctac	atgtcacgta	4080
4081	tgtgccatcc	caggagagga	acttcaccac	agcgccagca	atttgtcatg	aaggcaaagc	atacttccct	cgtgaaggtg	4160
4161	tttttgtgtt	taatggcact	tcttggttta	ttacacagag	gaacttcttt	tctccacaaa	taattactac	agacaataca	4240
4241	tttgtctcag	gaaattgtga	tgtcgttatt	ggcatcatta	acaacacagt	ttatgatcct	ctgcaacctg	agcttgactc	4320
4321	attcaaagaa	gagctggaca	agtacttcaa	aaatcataca	tcaccagatg	ttgatcttgg	cgacatttca	ggcattaacg	4400
4401	cttctgtcgt	caacattcaa	aaagaaattg	accgcctcaa	tgaggtcgct	aaaaatttaa	atgaatcact	cattgacctt	4480

```
4481 caagaattgg gaaaatatga gcaatatatt aaatggcctt ggtatgtttg gctcggcttc attgctggac taattgccat 4560
4561 cgtcatggtt acaatcttgc tttgttgcat gactagttgt tgcagttgcc tcaagggtgc atgctcttgt ggttcttgct 4640
4641 gcaagtttga tgaggatgac tctgagccag ttctcaaggg tgtcaaatta cattacacat aaGAATTCTG CAGATATCCA 4720
4721 GCACAGTGGC GGCCGCTCGA GTCTAGAGGG CCCGTTTAAA CCCGCTGATC AGCCTCGACT GTGCCTTCTA GTTGCCAGCC 4800
4801 ATCTGTTGTT TGCCCCTCCC CCGTGCCTTC CTTGACCCTG GAAGGTGCCA CTCCCACTGT CCTTTCCTAA TAAAATGAGG 4880
4881 AAATTGCATC GCATTGTCTG AGTAGGTGTC ATTCTATTCT GGGGGGTGGG GTGGGGCAGG ACAGCAAGGG GGAGGATTGG 4960
4961 GAAGACAATA GCAGGCATGC TGGGGATGCG GTGGGCTCTA TGGCTTCTGA GGCGGAAAGA ACCAGCTGGG GCTCTAGGGG 5040
5041 GTATCCCCAC GCGCCCTGTA GCGGCGCATT AAGCGCGGCG GGTGTGGTGG TTACGCGCAG CGTGACCGCT ACACTTGCCA 5120
5121 GCGCCCTAGC GCCCGCTCCT TTCGCTTTCT TCCCTTCCTT TCTCGCCACG TTCGCCGGCT TTCCCCGTCA AGCTCTAAAT 5200
5201 CGGGGGCTCC CTTTAGGGTT CCGATTTAGT GCTTTACGGC ACCTCGACCC CAAAAAACTT GATTAGGGTG ATGGTTCACG 5280
5281 TAGTGGGCCA TCGCCCTGAT AGACGGTTTT TCGCCCTTTG ACGTTGGAGT CCACGTTCTT TAATAGTGGA CTCTTGTTCC 5360
5361 AAACTGGAAC AACACTCAAC CCTATCTCGG TCTATTCTTT TGATTTATAA GGGATTTTGC CGATTTCGGC CTATTGGTTA 5440
5441 AAAAATGAGC TGATTTAACA AAAATTTAAC GCGAATTAAT TCTGTGGAAT GTGTGTCAGT TAGGGTGTGG AAAGTCCCCA 5520
5521 GGCTCCCCAG CAGGCAGAAG TATGCAAAGC ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAAGTCCC CAGGCTCCCC 5600
5601 AGCAGGCAGA AGTATGCAAA GCATGCATCT CAATTAGTCA GCAACCATAG TCCCGCCCCT AACTCCGCCC ATCCCGCCCC 5680
5681 TAACTCCGCC CAGTTCCGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA GGCCGCCTCT 5760
5761 GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG CTTTTGCAAA AAGCTCCCGG GAGCTTGTAT 5840
5841 ATCCATTTTC GGATCTGATC AAGAGACAGG ATGAGGATCG TTTCGCATGA TTGAACAAGA TGGATTGCAC GCAGGTTCTC 5920
5921 CGGCCGCTTG GGTGGAGAGG CTATTCGGCT ATGACTGGGC ACAACAGACA ATCGGCTGCT CTGATGCCGC CGTGTTCCGG 6000
6001 CTGTCAGCGC AGGGGCGCCC GGTTCTTTTT GTCAAGACCG ACCTGTCCGG TGCCCTGAAT GAACTGCAGG ACGAGGCAGC 6080
6081 GCGGCTATCG TGGCTGGCCA CGACGGGCGT TCCTTGCGCA GCTGTGCTCG ACGTTGTCAC TGAAGCGGGA AGGGACTGGC 6160
6161 TGCTATTGGG CGAAGTGCCG GGGCAGGATC TCCTGTCATC TCACCTTGCT CCTGCCGAGA AAGTATCCAT CATGGCTGAT 6240
6241 GCAATGCGGC GGCTGCATAC GCTTGATCCG GCTACCTGCC CATTCGACCA CCAAGCGAAA CATCGCATCG AGCGAGCACG 6320
6321 TACTCGGATG GAAGCCGGTC TTGTCGATCA GGATGATCTG GACGAAGAGC ATCAGGGGCT CGCGCCAGCC GAACTGTTCG 6400
6401 CCAGGCTCAA GGCGCGCATG CCCGACGGCG AGGATCTCGT CGTGACCCAT GGCGATGCCT GCTTGCCGAA TATCATGGTG 6480
6481 GAAAATGGCC GCTTTTCTGG ATTCATCGAC TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG CGTTGGCTAC 6560
6561 CCGTGATATT GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT CCCGATTCGC 6640
6641 AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG AGCGGGACTC TGGGGTTCGA AATGACCGAC CAAGCGACGC 6720
6721 CCAACCTGCC ATCACGAGAT TTCGATTCCA CCGCCGCCTT CTATGAAAGG TTGGGCTTCG GAATCGTTTT CCGGGACGCC 6800
6801 GGCTGGATGA TCCTCCAGCG CGGGGATCTC ATGCTGGAGT TCTTCGCCCA CCCCAACTTG TTTATTGCAG CTTATAATGG 6880
6881 TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC TAGTTGTGGT TTGTCCAAAC 6960
6961 TCATCAATGT ATCTTATCAT GTCTGTATAC CGTCGACCTC TAGCTAGAGC TTGGCGTAAT CATGGTCATA GCTGTTTCCT 7040
7041 GTGTGAAATT GTTATCCGCT CACAATTCCA CACAACATAC GAGCCGGAAG CATAAAGTGT AAAGCCTGGG GTGCCTAATG 7120
7121 AGTGAGCTAA CTCACATTAA TTGCGTTGCG CTCACTGCCC GCTTTCCAGT CGGGAAACCT GTCGTGCCAG CTGCATTAAT 7200
7201 GAATCGGCCA ACGCGCGGGG AGAGGCGGTT TGCGTATTGG GCGCTCTTCC GCTTCCTCGC TCACTGACTC GCTGCGCTCG 7280
7281 GTCGTTCGGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG 7360
7361 AAAGAACATG TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC CATAGGCTCC 7440
7441 GCCCCCCTGA CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG 7520
7521 TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC 7600
7601 GGGAAGCGTG GCGCTTTCTC ATAGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG 7680
7681 TGCACGAACC CCCCGTTCAG CCCGACCGCT GCGCCTTATC CGGTAACTAT CGTCTTGAGT CCAACCCGGT AAGACACGAC 7760
7761 TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA TGTAGGCGGT GCTACAGAGT TCTTGAAGTG 7840
7841 GTGGCCTAAC TACGGCTACA CTAGAAGAAC AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG 7920
7921 TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTTTTTTT GTTTGCAAGC AGCAGATTAC GCGCAGAAAA 8000
```

```
8001 AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT  CTGACGCTCA GTGGAACGAA AACTCACGTT AAGGGATTTT 8080
8081 GGTCATGAGA TTATCAAAAA GGATCTTCAC CTAGATCCTT  TTAAATTAAA AATGAAGTTT TAAATCAATC TAAAGTATAT 8160
8161 ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG  TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC 8240
8241 ATAGTTGCCT GACTCCCCGT CGTGTAGATA ACTACGATAC  GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC 8320
8321 GCGAGACCCA CGCTCACCGG CTCCAGATTT ATCAGCAATA  AACCAGCCAG CCGGAAGGGC CGAGCGCAGA AGTGGTCCTG 8400
8401 CAACTTTATC CGCCTCCATC CAGTCTATTA ATTGTTGCCG  GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC 8480
8481 AACGTTGTTG CCATTGCTAC AGGCATCGTG GTGTCACGCT  CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG 8560
8561 ATCAAGGCGA GTTACATGAT CCCCCATGTT GTGCAAAAAA  GCGGTTAGCT CCTTCGGTCC TCCGATCGTT GTCAGAAGTA 8640
8641 AGTTGGCCGC AGTGTTATCA CTCATGGTTA TGGCAGCACT  GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT 8720
8721 TCTGTGACTG GTGAGTACTC AACCAAGTCA TTCTGAGAAT  AGTGTATGCG GCGACCGAGT TGCTCTTGCC CGGCGTCAAT 8800
8801 ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG  CTCATCATTG GAAAACGTTC TTCGGGGCGA AAACTCTCAA 8880
8881 GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC  TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC 8960
8961 AGCGTTTCTG GGTGAGCAAA AACAGGAAGG CAAAATGCCG  CAAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT 9040
9041 CATACTCTTC CTTTTTCAAT ATTATTGAAG CATTTATCAG  GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT 9120
9121 AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG  AAAAGTGCCA CCTGACGTC                        9179
                |    10    |    20    |    30    |     40    |    50    |    60    |    70    |    80
```

pcDNA3-S1 comprises the first domain of the S (spike) protein (SEQ ID NO:26):

Vector pcDNA3.1(+) (UPPERCASE nt's)

S1: lower case/bold/<u>underscored</u>

```
     |    10    |    20    |    30    |    40     |    50    |    60    |    70    |    80    |
   1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT  CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGTAT 80
  81 CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG  CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA 160
 161 CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG  CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT 240
 241 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC  ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA 320
 321 CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC  CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT 400
 401 AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG  TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT 480
 481 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG  TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA 560
 561 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA  TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA 640
 641 TGGGCGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT  CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC 720
 721 AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC  CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG 800
 801 GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA  CTGCTTACTG GCTTATCGAA ATTAATACGA CTCACTATAG 880
 881 GGAGACCCAA GCTGGCTAGC GTTTAAACTT AAGCTTGGTA  CCGAGCTCGG ATCCatgttt attttcttat tatttcttac 960
 961 tctcactagt ggtagtgacc ttgaccggtg caccactttt  gatgatgttc aagctcctaa ttacactcaa catacttcat 1040
1041 ctatgagggg ggtttactat cctgatgaaa tttttagatc  agacactctt tatttaactc aggatttatt tcttccattt 1120
1121 tattctaatg ttacagggtt tcatactatt aatcatacgt  ttggcaaccc tgtcatacct tttaaggatg gtatttattt 1200
1201 tgctgccaca gagaaatcaa atgttgtccg tggttgggtt  tttggttcta ccatgaacaa caagtcacag tcggtgatta 1280
1281 ttattaacaa ttctactaat gttgttatac gagcatgtaa  ctttgaattg tgtgacaacc ctttctttgc tgtttctaaa 1360
1361 cccatgggta cacagacaca tactatgata ttcgataatg  catttaattg cactttcgag tacatatctg atgccttttc 1440
1441 gcttgatgtt tcagaaaagt caggtaattt taaacactta  cgagagtttg tgtttaaaaa taaagatggg tttctctatg 1520
```

```
1521 tttataaggg ctatcaacct atagatgtag ttcgtgatct  accttctggt tttaacactt tgaaacctat ttttaagttg 1600
1601 cctcttggta ttaacattac aaattttaga gccattctta  cagccttttc acctgctcaa gacatttggg gcacgtcagc 1680
1681 tgcagcctat tttgttggct atttaaagcc aactacattt  atgctcaagt atgatgaaaa tggtacaatc acagatgctg 1760
1761 ttgattgttc tcaaaatcca cttgctgaac tcaaatgctc  tgttaagagc tttgagattg acaaaggaat ttaccagacc 1840
1841 tctaatttca gggttgttcc ctcaggagat gttgtgagat  tccctaatat tacaaacttg tgtccttttg gagaggtttt 1920
1921 taatgctact aaattccctt ctgtctatgc atgggagaga  aaaaaaattt ctaattgtgt tgctgattac tctgtgctct 2000
2001 acaactcaac attttttca acctttaagt gctatggcgt  ttctgccact aagttgaatg atctttgctt ctccaatgtc 2080
2081 tatgcagatt cttttgtagt caagggagat gatgtaagac  aaatagcgcc aggacaaact ggtgttattg ctgattataa 2160
2161 ttataaattg ccagatgatt tcatgggttg tgtccttgct  tggaatacta ggaacattga tgctacttca actggtaatt 2240
2241 ataattataa atataggtat cttagacatg gcaagcttag  gccctttgag agagacatat ctaatgtgcc tttctcccct 2320
2321 gatggcaaac cttgcacccc acctgctctt aattgttatt  ggccattaaa tgattatggt tttacacca ctactggcat 2400
2401 tggctaccaa ccttacagag ttgtagtact ttcttttgaa  cttttaaatg caccggccac ggtttgtgga ccaaaattat 2480
2481 ccactgacct tattaagaac cagtgtgtca attttaattt  taatggactc actggtactg gtgtgttaac tccttcttca 2560
2561 aagagatttc aaccatttca acaatttggc cgtgatgttt  ctgatttcac tgattccgtt cgagatccta aaacatctga 2640
2641 aatattagac atttcacctt gctcttttgg gggtgtaagt  gtaattacac ctggaacaaa tgcttcatct gaagttgctg 2720
2721 ttctatatca agatgttaac tgcactgatg tttctacagc  aattcatgca gatcaactca caccagcttg gcgcatatat 2800
2801 tctactggaa acaatgtatt ccagactcaa gcaggctgtc  ttataggagc tgagcatgtc gacacttctt atgagtgcga 2880
2881 cattcctatt ggagctggca tttgtgctag ttaccataca  gtttctttat tacgtagtac tagccaaaaa tctattgtgg 2960
2961 cttatactat gtcttaaGAA TTCTGCAGAT ATCCAGCACA  GTGGCGGCCG CTCGAGTCTA GAGGGCCCGT TTAAACCCGC 3040
3041 TGATCAGCCT CGACTGTGCC TTCTAGTTGC CAGCCATCTG  TTGTTTGCCC CTCCCCCGTG CCTTCCTTGA CCCTGGAAGG 3120
3121 TGCCACTCCC ACTGTCCTTT CCTAATAAAA TGAGGAAATT  GCATCGCATT GTCTGAGTAG GTGTCATTCT ATTCTGGGGG 3200
3201 GTGGGGTGGG GCAGGACAGC AAGGGGGAGG ATTGGGAAGA  CAATAGCAGG CATGCTGGGG ATGCGGTGGG CTCTATGGCT 3280
3281 TCTGAGGCGG AAAGAACCAG CTGGGGCTCT AGGGGGTATC  CCCACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT 3360
3361 GGTGGTTACG CGCAGCGTGA CCGCTACACT TGCCAGCGCC  CTAGCGCCCG CTCCTTTCGC TTTCTTCCCT TCCTTTCTCG 3440
3441 CCACGTTCGC CGGCTTTCCC CGTCAAGCTC TAAATCGGGG  GCTCCCTTTA GGGTTCCGAT TTAGTGCTTT ACGGCACCTC 3520
3521 GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGTAGTG  GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT 3600
3601 GGAGTCCACG TTCTTTAATA GTGGACTCTT GTTCCAAACT  GGAACAACAC TCAACCCTAT CTCGGTCTAT TCTTTTGATT 3680
3681 TATAAGGGAT TTTGCCGATT TCGGCCTATT GGTTAAAAAA  TGAGCTGATT TAACAAAAAT TTAACGCGAA TTAATTCTGT 3760
3761 GGAATGTGTG TCAGTTAGGG TGTGGAAAGT CCCCAGGCTC  CCCAGCAGGC AGAAGTATGC AAAGCATGCA TCTCAATTAG 3840
3841 TCAGCAACCA GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG  GCAGAAGTAT GCAAAGCATG CATCTCAATT AGTCAGCAAC 3920
3921 CATAGTCCCG CCCCTAACTC CGCCCATCCC GCCCCTAACT  CCGCCCAGTT CCGCCCATTC TCCGCCCCAT GGCTGACTAA 4000
4001 TTTTTTTTAT TTATGCAGAG GCCGAGGCCG CCTCTGCCTC  TGAGCTATTC CAGAAGTAGT GAGGAGGCTT TTTTGGAGGC 4080
4081 CTAGGCTTTT GCAAAAAGCT CCCGGGAGCT TGTATATCCA  TTTTCGGATC TGATCAAGAG ACAGGATGAG GATCGTTTCG 4160
4161 CATGATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC  GCTTGGGTGG AGAGGCTATT CGGCTATGAC TGGGCACAAC 4240
4241 AGACAATCGG CTGCTCTGAT GCCGCCGTGT TCCGGCTGTC  AGCGCAGGGG CGCCCGGTTC TTTTTGTCAA GACCGACCTG 4320
4321 TCCGGTGCCC TGAATGAACT GCAGGACGAG GCAGCGCGGC  TATCGTGGCT GGCCACGACG GGCGTTCCTT GCGCAGCTGT 4400
4401 GCTCGACGTT GTCACTGAAG CGGGAAGGGA CTGGCTGCTA  TTGGGCGAAG TGCCGGGGCA GGATCTCCTG TCATCTCACC 4480
4481 TTGCTCCTGC CGAGAAAGTA TCCATCATGG CTGATGCAAT  GCGGCGGCTG CATACGCTTG ATCCGGCTAC CTGCCCATTC 4560
4561 GACCACCAAG CGAAACATCG CATCGAGCGA GCACGTACTC  GGATGGAAGC CGGTCTTGTC GATCAGGATG ATCTGGACGA 4640
4641 AGAGCATCAG GGGCTCGCGC CAGCCGAACT GTTCGCCAGG  CTCAAGGCGC GCATGCCCGA CGGCGAGGAT CTCGTCGTGA 4720
4721 CCCATGGCGA TGCCTGCTTG CCGAATATCA TGGTGGAAAA  TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT 4800
4801 GTGGCGGACC GCTATCAGGA CATAGCGTTG GCTACCCGTG  ATATTGCTGA AGAGCTTGGC GGCGAATGGG CTGACCGCTT 4880
4881 CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC  ATCGCCTTCT ATCGCCTTCT TGACGAGTTC TTCTGAGCGG 4960
4961 GACTCTGGGG TTCGAAATGA CCGACCAAGC GACGCCCAAC  CTGCCATCAC GAGATTTCGA TTCCACCGCC GCCTTCTATG 5040
```

```
5041 AAAGGTTGGG CTTCGGAATC GTTTTCCGGG ACGCCGGCTG GATGATCCTC CAGCGCGGGG ATCTCATGCT GGAGTTCTTC 5120
5121 GCCCACCCCA ACTTGTTTAT TGCAGCTTAT AATGGTTACA AATAAAGCAA TAGCATCACA AATTTCACAA ATAAAGCATT 5200
5201 TTTTTCACTG CATTCTAGTT GTGGTTTGTC CAAACTCATC AATGTATCTT ATCATGTCTG TATACCGTCG ACCTCTAGCT 5280
5281 AGAGCTTGGC GTAATCATGG TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC 5360
5361 GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC ATTAATTGCG TTGCGCTCAC TGCCCGCTTT 5440
5441 CCAGTCGGGA AACCTGTCGT GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT 5520
5521 CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA 5600
5601 ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA 5680
5681 AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT 5760
5761 GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG 5840
5841 CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA GGTATCTCAG 5920
5921 TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA 6000
6001 ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG 6080
6081 AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA AGAACAGTAT TTGGTATCTG 6160
6161 CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTT 6240
6241 TTTTTGTTTG CAAGCAGCAG ATTACGCGCA GAAAAAAAGG ATCTCAAGAA GATCCTTTGA TCTTTTCTAC GGGGTCTGAC 6320
6321 GCTCAGTGGA ACGAAAACTC ACGTTAAGGG ATTTTGGTCA TGAGATTATC AAAAAGGATC TTCACCTAGA TCCTTTTAAA 6400
6401 TTAAAAATGA AGTTTTAAAT CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA CCAATGCTTA ATCAGTGAGG 6480
6481 CACCTATCTC AGCGATCTGT CTATTTCGTT CATCCATAGT TGCCTGACTC CCCGTCGTGT AGATAACTAC GATACGGGAG 6560
6561 GGCTTACCAT CTGGCCCCAG TGCTGCAATG ATACCGCGAG ACCCACGCTC ACCGGCTCCA GATTTATCAG CAATAAACCA 6640
6641 GCCAGCCGGA AGGGCCGAGC GCAGAAGTGG TCCTGCAACT TTATCCGCCT CCATCCAGTC TATTAATTGT TGCCGGGAAG 6720
6721 CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT TGTTGCCATT GCTACAGGCA TCGTGGTGTC ACGCTCGTCG 6800
6801 TTTGGTATGG CTTCATTCAG CTCCGGTTCC CAACGATCAA GGCGAGTTAC ATGATCCCCC ATGTTGTGCA AAAAAGCGGT 6880
6881 TAGCTCCTTC GGTCCTCCGA TCGTTGTCAG AAGTAAGTTG GCCGCAGTGT TATCACTCAT GGTTATGGCA GCACTGCATA 6960
6961 ATTCTCTTAC TGTCATGCCA TCCGTAAGAT GCTTTTCTGT GACTGGTGAG TACTCAACCA AGTCATTCTG AGAATAGTGT 7040
7041 ATGCGGCGAC CGAGTTGCTC TTGCCCGGCG TCAATACGGG ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT 7120
7121 CATTGGAAAA CGTTCTTCGG GGCGAAAACT CTCAAGGATC TTACCGCTGT TGAGATCCAG TTCGATGTAA CCCACTCGTG 7200
7201 CACCCAACTG ATCTTCAGCA TCTTTTACTT TCACCAGCGT TTCTGGGTGA GCAAAAACAG GAAGGCAAAA TGCCGCAAAA 7280
7281 AAGGGAATAA GGGCGACACG GAAATGTTGA ATACTCATAC TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA 7360
7361 TTGTCTCATG AGCGGATACA TATTTGAATG TATTTAGAAA AATAAACAAA TAGGGGTTCC GCGCACATTT CCCCGAAAAG 7440
7441 TGCCACCTGA CGTC                                                                      7454
     |   10    |   20    |   30    |   40    |   50    |   60    |   70    |   80    |
```

pcDNA3-CRT/S1 construct comprising the human CRT sequence and S1 domain of the SARS-CoV S protein: (SEQ ID NO:27)

pcDNA3.1(+) vector (from Invitrogen) – sequence both 5' and 3' of the CRT and S1 sequences: UPPERCASE nt's

CRT sequence : *lower case/italic*

S1 sequence – **<u>lower case, bold/underscored</u>**

```
     |   10    |   20    |   30    |   40    |   50    |   60    |   70    |   80    |
  1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGTAT  80
 81 CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA 160
```

```
161 CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT 240
241 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA 320
321 CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT 400
401 AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT 480
481 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA 560
561 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA 640
641 TGGGCGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC 720
721 AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG 800
801 GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTACTG GCTTATCGAA ATTAATACGA CTCACTATAG 880
881 GGAGACCCAA GCTGGCTAGC GTTTAAACTT AAGatgctgc tccctgtgcc gctgctgctc ggcctgctcg gcctggccgc 960
961 cgccgagccc gtcgtctact tcaaggagca gtttctggac ggagatgggt ggaccgagcg ctggatcgaa tccaaacaca 1040
1041 agtccgattt tggcaaattc gtcctcagtt cgggcaagtt ctacggcgat caggagaaag ataaagggct gcagaccagc 1120
1121 caggacgccc gcttctacgc cctgtcggcc cgattcgagc cgttcagcaa caagggccag ccactggtgg tgcagttcac 1200
1201 cgtgaaacac gagcagaaca ttgactgcgg gggcggctac gtgaagctgt ttccggccgg cctggaccag aaggacatgc 1280
1281 acggggactc tgagtacaac atcatgtttg gtcctgacat ctgtggcccc ggcaccaaga aggttcacgt catcttcaac 1360
1361 tacaagggca agaacgtgct gatcaacaag gacatccgtt gcaaggacga cgagttcaca cacctgtaca cgctgatcgt 1440
1441 gcggccggac aacacgtatg aggtgaagat tgacaacagc caggtggagt cgggctccct ggaggatgac tgggacttcc 1520
1521 taccccccaa gaagataaag gacccagatg cctcgaagcc tgaagactgg gacgagcggg ccaagatcga cgaccccacg 1600
1601 gactccaagc ccgaggactg ggacaagccc gagcacatcc ccgacccgga cgcgaagaag cccgaagact gggacgaaga 1680
1681 aatggacgga gagtgggagc cgccggtgat tcagaacccc gagtacaagg gtgagtggaa gccgcggcag atcgacaacc 1760
1761 ccgattacaa aggcacctgg atccaccccg aaatcgacaa ccccgagtac tcgcccgacg ctaacatcta tgcctacgac 1840
1841 agctttgccg tgctgggctt ggacctctgg caggtcaagt cgggcaccat cttcgacaac ttcctcatca ccaacgatga 1920
1921 ggcgtacgca gaggagtttg gcaacgagac gtggggcgtc accaagacgg ccgagaagca gatgaaagac aagcaggacg 2000
2001 aggagcagcg gctgaaggag gaggaggagg agaagaagcg gaaggaggag gaggaggccg aggaggacga ggaggacaag 2080
2081 gacgacaagg aggacgagga tgaggacgag gaggacaagg acgaggagga ggaggaggcg gccgccggcc aggccaagga 2160
2161 cgagctgAGA TCCatgttta ttttcttatt atttcttact ctcactagtg gtagtgacct tgaccggtgc accacttttg 2240
2241 atgatgttca agctcctaat tacactcaac atacttcatc tatgaggggg gtttactatc ctgatgaaat tttagatca 2320
2321 gacactcttt atttaactca ggatttattt cttccatttt attctaatgt tacagggttt catactatta atcatacgtt 2400
2401 tggcaaccct gtcatacctt ttaaggatgg tatttatttt gctgccacag agaaatcaaa tgttgtccgt ggttgggttt 2480
2481 ttggttctac catgaacaac aagtcacagt cggtgattat tattaacaat tctactaatg ttgttatacg agcatgtaac 2560
2561 tttgaattgt gtgacaaccc tttctttgct gtttctaaac ccatgggtac acagacacat actatgatat tcgataatgc 2640
2641 atttaattgc actttcgagt acatatctga tgccttttcg cttgatgttt cagaaaagtc aggtaatttt aaacacttac 2720
2721 gagagtttgt gtttaaaaat aaagatgggt ttctctatgt ttataagggc tatcaaccta tagatgtagt tcgtgatcta 2800
2801 ccttctggtt ttaacacttt gaaacctatt tttaagttgc ctcttggtat taacattaca aattttagag ccattcttac 2880
2881 agccttttca cctgctcaag acatttgggg cacgtcagct gcagcctatt ttgttggcta tttaaagcca actacattta 2960
2961 tgctcaagta tgatgaaaat ggtacaatca cagatgctgt tgattgttct caaaatccac ttgctgaact caaatgctct 3040
3041 gttaagagct ttgagattga caaaggaatt taccagacct ctaatttcag ggttgttccc tcaggagatg ttgtgagatt 3120
3121 ccctaatatt acaaacttgt gtccttttgg agaggttttt aatgctacta aattcccttc tgtctatgca tgggagagaa 3200
3201 aaaaaatttc taattgtgtt gctgattact ctgtgctcta caactcaaca ttttttttcaa cctttaagtg ctatggcgtt 3280
3281 tctgccacta agttgaatga tctttgcttc tccaatgtct atgcagattc ttttgtagtc aagggagatg atgtaagaca 3360
3361 aatagcgcca ggacaaactg gtgttattgc tgattataat tataaattgc cagatgattt catgggttgt gtccttgctt 3440
3441 ggaatactag gaacattgat gctacttcaa ctggtaatta taattataaa tataggtatc ttagacatgg caagcttagg 3520
3521 ccctttgaga gagacatatc taatgtgcct ttctcccctg atggcaaacc ttgcaccccca cctgctctta attgttattg 3600
3601 gccattaaat gattatggtt tttacaccac tactggcatt ggctaccaac cttacagagt tgtagtactt tcttttgaac 3680
```

```
3681 ttttaaatgc accggccacg gtttgtggac caaaattatc cactgacctt attaagaacc agtgtgtcaa ttttaatttt 3760
3761 aatggactca ctggtactgg tgtgttaact ccttcttcaa agagatttca accatttcaa caatttggcc gtgatgtttc 3840
3841 tgatttcact gattccgttc gagatcctaa aacatctgaa atattagaca tttcaccttg ctcttttggg ggtgtaagtg 3920
3921 taattacacc tggaacaaat gcttcatctg aagttgctgt tctatatcaa gatgttaact gcactgatgt ttctacagca 4000
4001 attcatgcag atcaactcac accagcttgg cgcatatatt ctactggaaa caatgtattc cagactcaag caggctgtct 4080
4081 tataggagct gagcatgtcg acacttctta tgagtgcgac attcctattg gagctggcat ttgtgctagt taccatacag 4160
4161 tttctttatt acgtagtact agccaaaaat ctattgtggc ttatactatg tcttaaGAAT TCTGCAGATA TCCAGCACAG 4240
4241 TGGCGGCCGC TCGAGTCTAG AGGGCCCGTT TAAACCCGCT GATCAGCCTC GACTGTGCCT TCTAGTTGCC AGCCATCTGT 4320
4321 TGTTTGCCCC TCCCCCGTGC CTTCCTTGAC CCTGGAAGGT GCCACTCCCA CTGTCCTTTC CTAATAAAAT GAGGAAATTG 4400
4401 CATCGCATTG TCTGAGTAGG TGTCATTCTA TTCTGGGGGG TGGGGTGGGG CAGGACAGCA AGGGGGAGGA TTGGGAAGAC 4480
4481 AATAGCAGGC ATGCTGGGGA TGCGGTGGGC TCTATGGCTT CTGAGGCGGA AAGAACCAGC TGGGGCTCTA GGGGGTATCC 4560
4561 CCACGCGCCC TGTAGCGGCG CATTAAGCGC GGCGGGTGTG GTGGTTACGC GCAGCGTGAC CGCTACACTT GCCAGCGCCC 4640
4641 TAGCGCCCGC TCCTTTCGCT TTCTTCCCTT CCTTTCTCGC CACGTTCGCC GGCTTTCCCC GTCAAGCTCT AAATCGGGGG 4720
4721 CTCCCTTTAG GGTTCCGATT TAGTGCTTTA CGGCACCTCG ACCCCAAAAA ACTTGATTAG GGTGATGGTT CACGTAGTGG 4800
4801 GCCATCGCCC TGATAGACGG TTTTTCGCCC TTTGACGTTG GAGTCCACGT TCTTTAATAG TGGACTCTTG TTCCAAACTG 4880
4881 GAACAACACT CAACCCTATC TCGGTCTATT CTTTTGATTT ATAAGGGATT TTGCCGATTT CGGCCTATTG GTTAAAAAAT 4960
4961 GAGCTGATTT AACAAAAATT TAACGCGAAT TAATTCTGTG GAATGTGTGT CAGTTAGGGT GTGGAAAGTC CCCAGGCTCC 5040
5041 CCAGCAGGCA GAAGTATGCA AAGCATGCAT CTCAATTAGT CAGCAACCAG GTGTGGAAAG TCCCCAGGCT CCCCAGCAGG 5120
5121 CAGAAGTATG CAAAGCATGC ATCTCAATTA GTCAGCAACC ATAGTCCCGC CCCTAACTCC GCCCATCCCG CCCCTAACTC 5200
5201 CGCCCAGTTC CGCCCATTCT CCGCCCCATG GCTGACTAAT TTTTTTTATT TATGCAGAGG CCGAGGCCGC CTCTGCCTCT 5280
5281 GAGCTATTCC AGAAGTAGTG AGGAGGCTTT TTTGGAGGCC TAGGCTTTTG CAAAAAGCTC CCGGGAGCTT GTATATCCAT 5360
5361 TTTCGGATCT GATCAAGAGA CAGGATGAGG ATCGTTTCGC ATGATTGAAC AAGATGGATT GCACGCAGGT TCTCCGGCCG 5440
5441 CTTGGGTGGA GAGGCTATTC GGCTATGACT GGGCACAACA GACAATCGGC TGCTCTGATG CCGCCGTGTT CCGGCTGTCA 5520
5521 GCGCAGGGGC GCCCGGTTCT TTTTGTCAAG ACCGACCTGT CCGGTGCCCT GAATGAACTG CAGGACGAGG CAGCGCGGCT 5600
5601 ATCGTGGCTG GCCACGACGG GCGTTCCTTG CGCAGCTGTG CTCGACGTTG TCACTGAAGC GGGAAGGGAC TGGCTGCTAT 5680
5681 TGGGCGAAGT GCCGGGGCAG GATCTCCTGT CATCTCACCT TGCTCCTGCC GAGAAAGTAT CCATCATGGC TGATGCAATG 5760
5761 CGGCGGCTGC ATACGCTTGA TCCGGCTACC TGCCCATTCG ACCACCAAGC GAAACATCGC ATCGAGCGAG CACGTACTCG 5840
5841 GATGGAAGCC GGTCTTGTCG ATCAGGATGA TCTGGACGAA GAGCATCAGG GGCTCGCGCC AGCCGAACTG TTCGCCAGGC 5920
5921 TCAAGGCGCG CATGCCCGAC GGCGAGGATC TCGTCGTGAC CCATGGCGAT GCCTGCTTGC CGAATATCAT GGTGGAAAAT 6000
6001 GGCCGCTTTT CTGGATTCAT CGACTGTGGC CGGCTGGGTG TGGCGGACCG CTATCAGGAC ATAGCGTTGG CTACCCGTGA 6080
6081 TATTGCTGAA GAGCTTGGCG GCGAATGGGC TGACCGCTTC CTCGTGCTTT ACGGTATCGC CGCTCCCGAT TCGCAGCGCA 6160
6161 TCGCCTTCTA TCGCCTTCTT GACGAGTTCT TCTGAGCGGG ACTCTGGGGT TCGAAATGAC CGACCAAGCG ACGCCCAACC 6240
6241 TGCCATCACG AGATTTCGAT TCCACCGCCG CCTTCTATGA AAGGTTGGGC TTCGGAATCG TTTTCCGGGA CGCCGGCTGG 6320
6321 ATGATCCTCC AGCGCGGGGA TCTCATGCTG GAGTTCTTCG CCCACCCCAA CTTGTTTATT GCAGCTTATA ATGGTTACAA 6400
6401 ATAAAGCAAT AGCATCACAA ATTTCACAAA TAAAGCATTT TTTTCACTGC ATTCTAGTTG TGGTTTGTCC AAACTCATCA 6480
6481 ATGTATCTTA TCATGTCTGT ATACCGTCGA CCTCTAGCTA GAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA 6560
6561 AATTGTTATC CGCTCACAAT TCCACACAAC ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG 6640
6641 CTAACTCACA TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCAGCTGCAT TAATGAATCG 6720
6721 GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCTC TTCCGCTTCC TCGCTCACTG ACTCGCTGCG CTCGGTCGTT 6800
6801 CGGCTGCGGC GAGCGGTATC AGCTCACTCA AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAAGAA 6880
6881 CATGTGAGCA AAAGGCCAGC AAAAGGCCAG GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCCATAGG CTCCGCCCCC 6960
6961 CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG ACAGGACTAT AAAGATACCA GGCGTTTCCC 7040
7041 CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTCGGGAAG 7120
7121 CGTGGCGCTT TCTCATAGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC TGTGTGCACG 7200
```

```
7201 AACCCCCCGT TCAGCCCGAC CGCTGCGCCT TATCCGGTAA  CTATCGTCTT GAGTCCAACC CGGTAAGACA CGACTTATCG 7280
7281 CCACTGGCAG CAGCCACTGG TAACAGGATT AGCAGAGCGA  GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC 7360
7361 TAACTACGGC TACACTAGAA GAACAGTATT TGGTATCTGC  GCTCTGCTGA AGCCAGTTAC CTTCGGAAAA AGAGTTGGTA 7440
7441 GCTCTTGATC CGGCAAACAA ACCACCGCTG GTAGCGGTTT  TTTTGTTTGC AAGCAGCAGA TTACGCGCAG AAAAAAAGGA 7520
7521 TCTCAAGAAG ATCCTTTGAT CTTTTCTACG GGGTCTGACG  CTCAGTGGAA CGAAAACTCA CGTTAAGGGA TTTTGGTCAT 7600
7601 GAGATTATCA AAAAGGATCT TCACCTAGAT CCTTTTAAAT  TAAAAATGAA GTTTTAAATC AATCTAAAGT ATATATGAGT 7680
7681 AAACTTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC  ACCTATCTCA GCGATCTGTC TATTTCGTTC ATCCATAGTT 7760
7761 GCCTGACTCC CCGTCGTGTA GATAACTACG ATACGGGAGG  GCTTACCATC TGGCCCCAGT GCTGCAATGA TACCGCGAGA 7840
7841 CCCACGCTCA CCGGCTCCAG ATTTATCAGC AATAAACCAG  CCAGCCGGAA GGGCCGAGCG CAGAAGTGGT CCTGCAACTT 7920
7921 TATCCGCCTC CATCCAGTCT ATTAATTGTT GCCGGGAAGC  TAGAGTAAGT AGTTCGCCAG TTAATAGTTT GCGCAACGTT 8000
8001 GTTGCCATTG CTACAGGCAT CGTGGTGTCA CGCTCGTCGT  TTGGTATGGC TTCATTCAGC TCCGGTTCCC AACGATCAAG 8080
8081 GCGAGTTACA TGATCCCCCA TGTTGTGCAA AAAAGCGGTT  AGCTCCTTCG GTCCTCCGAT CGTTGTCAGA AGTAAGTTGG 8160
8161 CCGCAGTGTT ATCACTCATG GTTATGGCAG CACTGCATAA  TTCTCTTACT GTCATGCCAT CCGTAAGATG CTTTTCTGTG 8240
8241 ACTGGTGAGT ACTCAACCAA GTCATTCTGA GAATAGTGTA  TGCGGCGACC GAGTTGCTCT TGCCCGGCGT CAATACGGGA 8320
8321 TAATACCGCG CCACATAGCA GAACTTTAAA AGTGCTCATC  ATTGGAAAAC GTTCTTCGGG GCGAAAACTC TCAAGGATCT 8400
8401 TACCGCTGTT GAGATCCAGT TCGATGTAAC CCACTCGTGC  ACCCAACTGA TCTTCAGCAT CTTTTACTTT CACCAGCGTT 8480
8481 TCTGGGTGAG CAAAAACAGG AAGGCAAAAT GCCGCAAAAA  AGGGAATAAG GGCGACACGG AAATGTTGAA TACTCATACT 8560
8561 CTTCCTTTTT CAATATTATT GAAGCATTTA TCAGGGTTAT  TGTCTCATGA GCGGATACAT ATTTGAATGT ATTTAGAAAA 8640
8641 ATAAACAAAT AGGGGTTCCG CGCACATTTC CCCGAAAAGT  GCCACCTGAC GTC                                 8693
        |    10    |    20    |    30    |    40    |    50    |    60    |    70    |    80    |
```

pcDNA3-Si: (SEQ ID NO:28)

Vector pcDNA3.1(+) (UPPER CASE)

Si polypeptide coding sequence: **lower case/bold/underscored**

```
        |    10    |    20    |    30    |    40    |    50    |    60    |    70    |    80    |
   1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT  CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGTAT   80
  81 CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG  CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA  160
 161 CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG  CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT  240
 241 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC  ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA  320
 321 CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC  CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT  400
 401 AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG  TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT  480
 481 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG  TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA  560
 561 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA  TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA  640
 641 TGGGCGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT  CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC  720
 721 AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC  CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG  800
 801 GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA  CTGCTTACTG GCTTATCGAA ATTAATACGA CTCACTATAG  880
 881 GGAGACCCAA GCTGGCTAGC GTTTAAACTT AAGCTTGGTA  CCGAGCTCGG ATCCatgggt tgtgtccttg cttggaatac  960
 961 taggaacatt gatgctactt caactggtaa ttataattat  aaatataggt atcttagaca tggcaagctt aggccctttg 1040
1041 agagagacat atctaatgtg cctttctccc ctgatggcaa  accttgcacc ccacctgctc ttaattgtta ttggccatta 1120
1121 aatgattatg gtttttacac cactactggc attggctacc  aaccttacag agttgtagta ctttcttttg aacttttaaa 1200
```

```
1201 tgcaccggcc acggtttgtg gaccaaaatt atccactgac cttattaaga accagtgtgt caattttaat tttaatggac 1280
1281 tcactggtac tggtgtgtta actccttctt caaagagatt tcaaccattt caacaatttg gccgtgatgt ttctgatttc 1360
1361 actgattccg ttcgagatcc taaaacatct gaaatattag acatttcacc ttgctctttt gggggtgtaa gtgtaattac 1440
1441 acctggaaca aatgcttcat ctgaagttgc tgttctatat caagatgtta actgcactga tgtttctaca gcaattcatg 1520
1521 cagatcaact cacaccagct tggcgcatat attctactgg aaacaatgta ttccagactc aagcaggctg tcttatagga 1600
1601 gctgagcatg tcgacacttc ttatgagtgc gacattccta ttggagctgg catttgtgct agttaccata cagtttcttt 1680
1681 attacgtagt actagccaaa aatctattgt ggcttatact atgtctttag gtgctgatag ttcaattgct tactctaata 1760
1761 acaccattgc tatacctact aacttttcaa ttagcattac tacagaagta atgcctgttt ctatggctaa aacctccgta 1840
1841 gattgtaata tgtacatctg cggagattct actgaatgtg ctaatttgct tctccaatat ggtagctttt gcacacaact 1920
1921 aaatcgtgca ctctcaggta ttgctgctga acaggatcgc aacacacgtg aagtgttcgc tcaagtcaaa caaatgtaca 2000
2001 aaaccccaac tttgaaatat tttggtggtt ttaatttttc acaaatatta cctgaccctc taaagccaac taagaggtct 2080
2081 tttattgagg acttgctctt taataaggtg acactcgctg atgctggctt catgtaaGAA TTCTGCAGAT ATCCAGCACA 2160
2161 GTGGCGGCCG CTCGAGTCTA GAGGGCCCGT TTAAACCCGC TGATCAGCCT CGACTGTGCC TTCTAGTTGC CAGCCATCTG 2240
2241 TTGTTTGCCC CTCCCCCGTG CCTTCCTTGA CCCTGGAAGG TGCCACTCCC ACTGTCCTTT CCTAATAAAA TGAGGAAATT 2320
2321 GCATCGCATT GTCTGAGTAG GTGTCATTCT ATTCTGGGGG GTGGGGTGGG GCAGGACAGC AAGGGGGAGG ATTGGGAAGA 2400
2401 CAATAGCAGG CATGCTGGGG ATGCGGTGGG CTCTATGGCT TCTGAGGCGG AAAGAACCAG CTGGGGCTCT AGGGGGTATC 2480
2481 CCCACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG CGCAGCGTGA CCGCTACACT TGCCAGCGCC 2560
2561 CTAGCGCCCG CTCCTTTCGC TTTCTTCCCT TCCTTTCTCG CCACGTTCGC CGGCTTTCCC CGTCAAGCTC TAAATCGGGG 2640
2641 GCTCCCTTTA GGGTTCCGAT TTAGTGCTTT ACGGCACCTC GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGTAGTG 2720
2721 GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA GTGGACTCTT GTTCCAAACT 2800
2801 GGAACAACAC TCAACCCTAT CTCGGTCTAT TCTTTTGATT TATAAGGGAT TTTGCCGATT TCGGCCTATT GGTTAAAAAA 2880
2881 TGAGCTGATT TAACAAAAAT TTAACGCGAA TTAATTCTGT GGAATGTGTG TCAGTTAGGG TGTGGAAAGT CCCCAGGCTC 2960
2961 CCCAGCAGGC AGAAGTATGC AAAGCATGCA TCTCAATTAG TCAGCAACCA GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG 3040
3041 GCAGAAGTAT GCAAAGCATG CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC GCCCCTAACT 3120
3121 CCGCCCAGTT CCGCCCATTC TCCGCCCCAT GGCTGACTAA TTTTTTTTAT TTATGCAGAG GCCGAGGCCG CCTCTGCCTC 3200
3201 TGAGCTATTC CAGAAGTAGT GAGGAGGCTT TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT CCCGGGAGCT TGTATATCCA 3280
3281 TTTTCGGATC TGATCAAGAG ACAGGATGAG GATCGTTTCG CATGATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC 3360
3361 GCTTGGGTGG AGAGGCTATT CGGCTATGAC TGGGCACAAC AGACAATCGG CTGCTCTGAT GCCGCCGTGT TCCGGCTGTC 3440
3441 AGCGCAGGGG CGCCCGGTTC TTTTTGTCAA GACCGACCTG TCCGGTGCCC TGAATGAACT GCAGGACGAG GCAGCGCGGC 3520
3521 TATCGTGGCT GGCCACGACG GGCGTTCCTT GCGCAGCTGT GCTCGACGTT GTCACTGAAG CGGGAAGGGA CTGGCTGCTA 3600
3601 TTGGGCGAAG TGCCGGGGCA GGATCTCCTG TCATCTCACC TTGCTCCTGC CGAGAAAGTA TCCATCATGG CTGATGCAAT 3680
3681 GCGGCGGCTG CATACGCTTG ATCCGGCTAC CTGCCCATTC GACCACCAAG CGAAACATCG CATCGAGCGA GCACGTACTC 3760
3761 GGATGGAAGC CGGTCTTGTC GATCAGGATG ATCTGGACGA AGAGCATCAG GGGCTCGCGC CAGCCGAACT GTTCGCCAGG 3840
3841 CTCAAGGCGC GCATGCCCGA CGGCGAGGAT CTCGTCGTGA CCCATGGCGA TGCCTGCTTG CCGAATATCA TGGTGGAAAA 3920
3921 TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT GTGGCGGACC GCTATCAGGA CATAGCGTTG GCTACCCGTG 4000
4001 ATATTGCTGA AGAGCTTGGC GGCGAATGGG CTGACCGCTT CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC 4080
4081 ATCGCCTTCT ATCGCCTTCT TGACGAGTTC TTCTGAGCGG GACTCTGGGG TTCGAAATGA CCGACCAAGC GACGCCCAAC 4160
4161 CTGCCATCAC GAGATTTCGA TTCCACCGCC GCCTTCTATG AAAGGTTGGG CTTCGGAATC GTTTTCCGGG ACGCCGGCTG 4240
4241 GATGATCCTC CAGCGCGGGG ATCTCATGCT GGAGTTCTTC GCCCACCCCA ACTTGTTTAT TGCAGCTTAT AATGGTTACA 4320
4321 AATAAAGCAA TAGCATCACA AATTTCACAA ATAAAGCATT TTTTTCACTG CATTCTAGTT GTGGTTTGTC CAAACTCATC 4400
4401 AATGTATCTT ATCATGTCTG TATACCGTCG ACCTCTAGCT AGAGCTTGGC GTAATCATGG TCATAGCTGT TTCCTGTGTG 4480
4481 AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA 4560
4561 GCTAACTCAC ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA TTAATGAATC 4640
4641 GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT CTTCCGCTTC CTCGCTCACT GACTCGCTGC GCTCGGTCGT 4720
```

```
4721 TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA  ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA 4800
4801 ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA  AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC 4880
4881 CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT  GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC 4960
4961 CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG  CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA 5040
5041 GCGTGGCGCT TTCTCATAGC TCACGCTGTA GGTATCTCAG  TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC 5120
5121 GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA  ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC 5200
5201 GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG  AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC 5280
5281 CTAACTACGG CTACACTAGA AGAACAGTAT TTGGTATCTG  CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT 5360
5361 AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTT  TTTTTGTTTG CAAGCAGCAG ATTACGCGCA GAAAAAAAGG 5440
5441 ATCTCAAGAA GATCCTTTGA TCTTTTCTAC GGGGTCTGAC  GCTCAGTGGA ACGAAAACTC ACGTTAAGGG ATTTTGGTCA 5520
5521 TGAGATTATC AAAAAGGATC TTCACCTAGA TCCTTTTAAA  TTAAAAATGA AGTTTTAAAT CAATCTAAAG TATATATGAG 5600
5601 TAAACTTGGT CTGACAGTTA CCAATGCTTA ATCAGTGAGG  CACCTATCTC AGCGATCTGT CTATTTCGTT CATCCATAGT 5680
5681 TGCCTGACTC CCCGTCGTGT AGATAACTAC GATACGGGAG  GGCTTACCAT CTGGCCCCAG TGCTGCAATG ATACCGCGAG 5760
5761 ACCCACGCTC ACCGGCTCCA GATTTATCAG CAATAAACCA  GCCAGCCGGA AGGGCCGAGC GCAGAAGTGG TCCTGCAACT 5840
5841 TTATCCGCCT CCATCCAGTC TATTAATTGT TGCCGGGAAG  CTAGAGTAAG TAGTTCGCCA GTTAATAGTT TGCGCAACGT 5920
5921 TGTTGCCATT GCTACAGGCA TCGTGGTGTC ACGCTCGTCG  TTTGGTATGG CTTCATTCAG CTCCGGTTCC CAACGATCAA 6000
6001 GGCGAGTTAC ATGATCCCCC ATGTTGTGCA AAAAAGCGGT  TAGCTCCTTC GGTCCTCCGA TCGTTGTCAG AAGTAAGTTG 6080
6081 GCCGCAGTGT TATCACTCAT GGTTATGGCA GCACTGCATA  ATTCTCTTAC TGTCATGCCA TCCGTAAGAT GCTTTTCTGT 6160
6161 GACTGGTGAG TACTCAACCA AGTCATTCTG AGAATAGTGT  ATGCGGCGAC CGAGTTGCTC TTGCCCGGCG TCAATACGGG 6240
6241 ATAATACCGC GCCACATAGC AGAACTTTAA AAGTGCTCAT  CATTGGAAAA CGTTCTTCGG GGCGAAAACT CTCAAGGATC 6320
6321 TTACCGCTGT TGAGATCCAG TTCGATGTAA CCCACTCGTG  CACCCAACTG ATCTTCAGCA TCTTTTACTT TCACCAGCGT 6400
6401 TTCTGGGTGA GCAAAAACAG GAAGGCAAAA TGCCGCAAAA  AAGGGAATAA GGGCGACACG GAAATGTTGA ATACTCATAC 6480
6481 TCTTCCTTTT TCAATATTAT TGAAGCATTT ATCAGGGTTA  TTGTCTCATG AGCGGATACA TATTTGAATG TATTTAGAAA 6560
6561 AATAAACAAA TAGGGGTTCC GCGCACATTT CCCCGAAAAG  TGCCACCTGA CGTC                                   6614
          |    10    |    20    |    30    |    40     |    50    |    60    |    70    |    80    |
```

pcDNA3-S2 (SEQ ID NO:29):

vector pcDNA3.1(+) sequence (UPPER CASE)
S2 – C-terminal domain of SARS-CoV S protein (**lower case/bold/underscored**)

```
          |    10    |    20    |    30    |    40     |    50    |    60    |    70    |    80    |
   1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT  CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGTAT   80
  81 CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG  CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA  160
 161 CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG  CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT  240
 241 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC  ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA  320
 321 CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC  CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT  400
 401 AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG  TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT  480
 481 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG  TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA  560
 561 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA  TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA  640
 641 TGGGCGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT  CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC  720
```

```
721 AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG 800
801 GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTACTG GCTTATCGAA ATTAATACGA CTCACTATAG 880
881 GGAGACCCAA GCTGGCTAGC GTTTAAACTT AAGCTTGGTA CCGAGCTCGG ATCCatgatg ttaggtgctg atagttcaat 960
961 tgcttactct aataacacca ttgctatacc tactaacttt tcaattagca ttactacaga agtaatgcct gtttctatgg 1040
1041 ctaaaacctc cgtagattgt aatatgtaca tctgcggaga ttctactgaa tgtgctaatt tgcttctcca atatggtagc 1120
1121 ttttgcacac aactaaatcg tgcactctca ggtattgctg ctgaacagga tcgcaacaca cgtgaagtgt tcgctcaagt 1200
1201 caaacaaatg tacaaaaccc caactttgaa atattttggt ggttttaatt tttcacaaat attacctgac cctctaaagc 1280
1281 caactaagag gtcttttatt gaggacttgc tctttaataa ggtgacactc gctgatgctg gcttcatgaa gcaatatggc 1360
1361 gaatgcctag gtgatattaa tgctagagat ctcatttgtg cgcagaagtt caatggactt acagtgttgc cacctctgct 1440
1441 cactgatgat atgattgctg cctacactgc tgctctagtt agtggtactg ccactgctgg atggacattt ggtgctggcg 1520
1521 ctgctcttca aatacctttt gctatgcaaa tggcatatag gttcaatggc attggagtta cccaaaatgt tctctatgag 1600
1601 aaccaaaaac aaatcgccaa ccaatttaac aaggcgatta gtcaaattca agaatcactt acaacaacat caactgcatt 1680
1681 gggcaagctg caagacgttg ttaaccagaa tgctcaagca ttaaacacac ttgttaaaca acttagctct aattttggtg 1760
1761 caatttcaag tgtgctaaat gatatccttt cgcgacttga taaagtcgag gcggaggtac aaattgacag gttaattaca 1840
1841 ggcagacttc aaagccttca aacctatgta acacaacaac taatcagggc tgctgaaatc agggcttctg ctaatcttgc 1920
1921 tgctactaaa atgtctgagt gtgttcttgg acaatcaaaa agagttgact tttgtggaaa gggctaccac cttatgtcct 2000
2001 tcccacaagc agccccgcat ggtgttgtct tcctacatgt cacgtatgtg ccatcccagg agaggaactt caccacagcg 2080
2081 ccagcaattt gtcatgaagg caaagcatac ttccctcgtg aaggtgtttt tgtgtttaat ggcacttctt ggtttattac 2160
2161 acagaggaac ttcttttctc cacaaataat tactacagac aatacatttg tctcaggaaa ttgtgatgtc gttattggca 2240
2241 tcattaacaa cacagtttat gatcctctgc aacctgagct tgactcattc aaagaagagc tggacaagta cttcaaaaat 2320
2321 catacatcac cagatgttga tcttggcgac atttcaggca ttaacgcttc tgtcgtcaac attcaaaaag aaattgaccg 2400
2401 cctcaatgag gtcgctaaaa atttaaatga atcactcatt gaccttcaag aattgggaaa atatgagcaa tatattaaat 2480
2481 ggccttggta tgtttggctc ggcttcattg ctggactaat tgccatcgtc atggttacaa tcttgctttg ttgcatgact 2560
2561 agttgttgca gttgcctcaa gggtgcatgc tcttgtggtt cttgctgcaa gtttgatgag gatgactctg agccagttct 2640
2641 caagggtgtc aaattacatt acacataaGA ATTCTGCAGA TATCCAGCAC AGTGGCGGCC GCTCGAGTCT AGAGGGCCCG 2720
2721 TTTAAACCCG CTGATCAGCC TCGACTGTGC CTTCTAGTTG CCAGCCATCT GTTGTTTGCC CCTCCCCCGT GCCTTCCTTG 2800
2801 ACCCTGGAAG GTGCCACTCC CACTGTCCTT TCCTAATAAA ATGAGGAAAT TGCATCGCAT TGTCTGAGTA GGTGTCATTC 2880
2881 TATTCTGGGG GGTGGGGTGG GGCAGGACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG 2960
2961 GCTCTATGGC TTCTGAGGCG GAAAGAACCA GCTGGGGCTC TAGGGGGTAT CCCCACGCGC CCTGTAGCGG CGCATTAAGC 3040
3041 GCGGCGGGTG TGGTGGTTAC GCGCAGCGTG ACCGCTACAC TTGCCAGCGC CCTAGCGCCC GCTCCTTTCG CTTTCTTCCC 3120
3121 TTCCTTTCTC GCCACGTTCG CCGGCTTTCC CCGTCAAGCT CTAAATCGGG GGCTCCCTTT AGGGTTCCGA TTTAGTGCTT 3200
3201 TACGGCACCT CGACCCCAAA AAACTTGATT AGGGTGATGG TTCACGTAGT GGGCCATCGC CCTGATAGAC GGTTTTTCGC 3280
3281 CCTTTGACGT TGGAGTCCAC GTTCTTTAAT AGTGGACTCT TGTTCCAAAC TGGAACAACA CTCAACCCTA TCTCGGTCTA 3360
3361 TTCTTTTGAT TTATAAGGGA TTTTGCCGAT TTCGGCCTAT TGGTTAAAAA ATGAGCTGAT TTAACAAAAA TTTAACGCGA 3440
3441 ATTAATTCTG TGGAATGTGT GTCAGTTAGG GTGTGGAAAG TCCCCAGGCT CCCCAGCAGG CAGAAGTATG CAAAGCATGC 3520
3521 ATCTCAATTA GTCAGCAACC AGGTGTGGAA AGTCCCCAGG CTCCCCAGCA GGCAGAAGTA TGCAAAGCAT GCATCTCAAT 3600
3601 TAGTCAGCAA CCATAGTCCC GCCCCTAACT CCGCCCATCC CGCCCCTAAC TCCGCCCAGT TCCGCCCATT CTCCGCCCCA 3680
3681 TGGCTGACTA ATTTTTTTTA TTTATGCAGA GGCCGAGGCC GCCTCTGCCT CTGAGCTATT CCAGAAGTAG TGAGGAGGCT 3760
3761 TTTTTGGAGG CCTAGGCTTT TGCAAAAAGC TCCCGGGAGC TTGTATATCC ATTTTCGGAT CTGATCAAGA GACAGGATGA 3840
3841 GGATCGTTTC GCATGATTGA ACAAGATGGA TTGCACGCAG GTTCTCCGGC CGCTTGGGTG GAGAGGCTAT TCGGCTATGA 3920
3921 CTGGGCACAA CAGACAATCG GCTGCTCTGA TGCCGCCGTG TTCCGGCTGT CAGCGCAGGG GCGCCCGGTT CTTTTTGTCA 4000
4001 AGACCGACCT GTCCGGTGCC CTGAATGAAC TGCAGGACGA GGCAGCGCGG CTATCGTGGC TGGCCACGAC GGGCGTTCCT 4080
4081 TGCGCAGCTG TGCTCGACGT TGTCACTGAA GCGGGAAGGG ACTGGCTGCT ATTGGGCGAA GTGCCGGGGC AGGATCTCCT 4160
4161 GTCATCTCAC CTTGCTCCTG CCGAGAAAGT ATCCATCATG GCTGATGCAA TGCGGCGGCT GCATACGCTT GATCCGGCTA 4240
```

```
4241 CCTGCCCATT CGACCACCAA GCGAAACATC GCATCGAGCG AGCACGTACT CGGATGGAAG CCGGTCTTGT CGATCAGGAT 4320
4321 GATCTGGACG AAGAGCATCA GGGGCTCGCG CCAGCCGAAC TGTTCGCCAG GCTCAAGGCG CGCATGCCCG ACGGCGAGGA 4400
4401 TCTCGTCGTG ACCCATGGCG ATGCCTGCTT GCCGAATATC ATGGTGGAAA ATGGCCGCTT TTCTGGATTC ATCGACTGTG 4480
4481 GCCGGCTGGG TGTGGCGGAC CGCTATCAGG ACATAGCGTT GGCTACCCGT GATATTGCTG AAGAGCTTGG CGGCGAATGG 4560
4561 GCTGACCGCT TCCTCGTGCT TTACGGTATC GCCGCTCCCG ATTCGCAGCG CATCGCCTTC TATCGCCTTC TTGACGAGTT 4640
4641 CTTCTGAGCG GGACTCTGGG GTTCGAAATG ACCGACCAAG CGACGCCCAA CCTGCCATCA CGAGATTTCG ATTCCACCGC 4720
4721 CGCCTTCTAT GAAAGGTTGG GCTTCGGAAT CGTTTTCCGG GACGCCGGCT GGATGATCCT CCAGCGCGGG GATCTCATGC 4800
4801 TGGAGTTCTT CGCCCACCCC AACTTGTTTA TTGCAGCTTA TAATGGTTAC AAATAAAGCA ATAGCATCAC AAATTTCACA 4880
4881 AATAAAGCAT TTTTTTCACT GCATTCTAGT TGTGGTTTGT CCAAACTCAT CAATGTATCT TATCATGTCT GTATACCGTC 4960
4961 GACCTCTAGC TAGAGCTTGG CGTAATCATG GTCATAGCTG TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA 5040
5041 ACATACGAGC CGGAAGCATA AAGTGTAAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGCGCTCA 5120
5121 CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT CGGCCAACGC GCGGGGAGAG GCGGTTTGCG 5200
5201 TATTGGGCGC TCTTCCGCTT CCTCGCTCAC TGACTCGCTG CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT 5280
5281 CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC 5360
5361 AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC 5440
5441 AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG 5520
5521 TTCCGACCCT GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCATAG CTCACGCTGT 5600
5601 AGGTATCTCA GTTCGGTGTA GGTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC 5680
5681 CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA 5760
5761 TTAGCAGAGC GAGGTATGTA GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGAACAGTA 5840
5841 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA TCCGGCAAAC AAACCACCGC 5920
5921 TGGTAGCGGT TTTTTTGTTT GCAAGCAGCA GATTACGCGC AGAAAAAAAG GATCTCAAGA AGATCCTTTG ATCTTTTCTA 6000
6001 CGGGGTCTGA CGCTCAGTGG AACGAAAACT CACGTTAAGG GATTTTGGTC ATGAGATTAT CAAAAAGGAT CTTCACCTAG 6080
6081 ATCCTTTTAA ATTAAAAATG AAGTTTTAAA TCAATCTAAA GTATATATGA GTAAACTTGG TCTGACAGTT ACCAATGCTT 6160
6161 AATCAGTGAG GCACCTATCT CAGCGATCTG TCTATTTCGT TCATCCATAG TTGCCTGACT CCCCGTCGTG TAGATAACTA 6240
6241 CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA GACCCACGCT CACCGGCTCC AGATTTATCA 6320
6321 GCAATAAACC AGCCAGCCGG AAGGGCCGAG CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT CTATTAATTG 6400
6401 TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT 6480
6481 CACGCTCGTC GTTTGGTATG GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA CATGATCCCC CATGTTGTGC 6560
6561 AAAAAAGCGG TTAGCTCCTT CGGTCCTCCG ATCGTTGTCA GAAGTAAGTT GGCCGCAGTG TTATCACTCA TGGTTATGGC 6640
6641 AGCACTGCAT AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG TGACTGGTGA GTACTCAACC AAGTCATTCT 6720
6721 GAGAATAGTG TATGCGGCGA CCGAGTTGCT CTTGCCCGGC GTCAATACGG GATAATACCG CGCCACATAG CAGAACTTTA 6800
6801 AAAGTGCTCA TCATTGGAAA ACGTTCTTCG GGGCGAAAAC TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA 6880
6881 ACCCACTCGT GCACCCAACT GATCTTCAGC ATCTTTTACT TTCACCAGCG TTTCTGGGTG AGCAAAAACA GGAAGGCAAA 6960
6961 ATGCCGCAAA AAAGGGAATA AGGGCGACAC GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT 7040
7041 TATCAGGGTT ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA ATAGGGGTTC CGCGCACATT 7120
7121 TCCCCGAAAA GTGCCACCTG ACGTC                                                               7145
      |   10    |   20    |   30    |   40    |   50    |   60    |   70    |   80    |
```

pcDNA3-CRT/M (SEQ ID NO:30)

Vector sequence pcDNA3.1 (-)mycHisA (UPPERCASE)
CRT Sequence *(lower case/italic)*
Sequence of M protein (lower case/**bold**/underscored)

```
        |   10     |   20     |   30     |   40      |   50     |   60     |   70     |   80
   1 GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCACTCT  CAGTACAATC TGCTCTGATG CCGCATAGTT AAGCCAGTAT 80
  81 CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG  CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA 160
 161 CAATTGCATG AAGAATCTGC TTAGGGTTAG GCGTTTTGCG  CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT 240
 241 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC  ATTAGTTCAT AGCCCATATA TGGAGTTCCG CGTTACATAA 320
 321 CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC  CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT 400
 401 AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAG  TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT 480
 481 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG  TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA 560
 561 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA  TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA 640
 641 TGGGCGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT  CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC 720
 721 AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC  CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG 800
 801 GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA  CTGCTTACTG GCTTATCGAA ATTAATACGA CTCACTATAG 880
 881 GGAGACCCAA GCTGGCTAGC GTTTAAACGG GCCCTCTAGA  atgctgctcc ctgtgccgct gctgctcggc ctgctcggcc 960
 961 tggccgccgc cgagcccgtc gtctacttca aggagcagtt  tctggacgga gatgggtgga ccgagcgctg gatcgaatcc 1040
1041 aaacacaagt ccgattttgg caaattcgtc ctcagttcgg  gcaagttcta cggcgatcag gagaaagata aagggctgca 1120
1121 gaccagccag gacgcccgct tctacgccct gtcggcccga  ttcgagccgt tcagcaacaa gggccagcca ctggtggtgc 1200
1201 agttcaccgt gaaacacgag cagaacattg actgcggggg  cggctacgtg aagctgtttc cggccggcct ggaccagaag 1280
1281 gacatgcacg gggactctga gtacaacatc atgtttggtc  ctgacatctg tggccccggc accaagaagg ttcacgtcat 1360
1361 cttcaactac aagggcaaga acgtgctgat caacaaggac  atccgttgca aggacgacga gttcacacac ctgtacacgc 1440
1441 tgatcgtgcg gccggacaac acgtatgagg tgaagattga  caacagccag gtggagtcgg gctccctgga ggatgactgg 1520
1521 gacttcctac cccccaagaa gataaaggac ccagatgcct  cgaagcctga agactgggac gagcgggcca agatcgacga 1600
1601 ccccacggac tccaagcccg aggactggga caagcccgag  cacatccccg acccggacgc gaagaagccc gaagactggg 1680
1681 acgaagaaat ggacggagag tgggagccgc cggtgattca  gaaccccgag tacaagggtg agtggaagcc gcggcagatc 1760
1761 gacaaccccg attacaaagg cacctggatc caccccgaaa  tcgacaaccc cgagtactcg cccgacgcta acatctatgc 1840
1841 ctacgacagc tttgccgtgc tgggcttgga cctctggcag  gtcaagtcgg gcaccatctt cgacaacttc ctcatcacca 1920
1921 acgatgaggc gtacgcagag gagtttggca acgagacgtg  ggcgtcaccc aagacggccg agaagcagat gaaagacaag 2000
2001 caggacgagg agcagcggct gaaggaggag gaggaggaga  agaagcggaa ggaggaggag gaggccgagg aggacgagga 2080
2081 ggacaaggac gacaaggagg acgaggatga ggacgaggag  gacaaggacg aggaggagga ggaggcggcc gccggccagg 2160
2161 ccaaggacga gctgtagGAA TTCatggcag acaacggtac  tattaccgtt gaggagctta aacaactcct ggaacaatgg 2240
2241 aacctagtaa taggtttcct attcctagcc tggattatgt  tactacaatt tgcctattct aatcggaaca ggtttttgta 2320
2321 cataataaag cttgttttcc tctggctctt gtggccagta  acacttgctt gttttgtgct tgctgctgtc tacagaatta 2400
2401 attgggtgac tggcgggatt gcgattgcaa tggcttgtat  tgtaggcttg atgtggctta gctacttcgt tgcttccttc 2480
2481 aggctgtttg ctcgtacccg ctcaatgtgg tcattcaacc  cagaaacaaa cattcttctc aatgtgcctc tccgggggac 2560
2561 aattgtgacc agaccgctca tggaaagtga acttgtcatt  ggtgctgtga tcattcgtgg tcacttgcga atggccggac 2640
2641 actccctagg gcgctgtgac attaaggacc tgccaaaaga  gatcactgtg gctacatcac gaacgctttc ttattacaaa 2720
2721 ttaggagcgt cgcagcgtgt aggcactgat tcaggttttg  ctgcatacaa ccgctaccgt attggaaact ataaattaaa 2800
2801 tacagaccac gccggtagca acgacaatat tgctttgcta  gtacagGGTA CCAAGCTTGG GCCCGAACAA AAACTCATCT 2880
2881 CAGAAGAGGA TCTGAATAGC GCCGTCGACC ATCATCATCA  TCATCATTGA GTTTAAACGG TCTCCAGCTT AAGTTTAAAC 2960
```

```
2961 CGCTGATCAG CCTCGACTGT GCCTTCTAGT TGCCAGCCAT CTGTTGTTTG CCCCTCCCCC GTGCCTTCCT TGACCCTGGA 3040
3041 AGGTGCCACT CCCACTGTCC TTTCCTAATA AAATGAGGAA ATTGCATCGC ATTGTCTGAG TAGGTGTCAT TCTATTCTGG 3120
3121 GGGGTGGGGT GGGGCAGGAC AGCAAGGGGG AGGATTGGGA AGACAATAGC AGGCATGCTG GGGATGCGGT GGGCTCTATG 3200
3201 GCTTCTGAGG CGGAAAGAAC CAGCTGGGGC TCTAGGGGGT ATCCCCACGC GCCCTGTAGC GGCGCATTAA GCGCGGCGGG 3280
3281 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT CGCTTTCTTC CCTTCCTTTC 3360
3361 TCGCCACGTT CGCCGGCTTT CCCCGTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC GATTTAGTGC TTTACGGCAC 3440
3441 CTCGACCCCA AAAAACTTGA TTAGGGTGAT GGTTCACGTA GTGGGCCATC GCCCTGATAG ACGGTTTTTC GCCCTTTGAC 3520
3521 GTTGGAGTCC ACGTTCTTTA ATAGTGGACT CTTGTTCCAA ACTGGAACAA CACTCAACCC TATCTCGGTC TATTCTTTTG 3600
3601 ATTTATAAGG GATTTTGCCG ATTTCGGCCT ATTGGTTAAA AAATGAGCTG ATTTAACAAA AATTTAACGC GAATTAATTC 3680
3681 TGTGGAATGT GTGTCAGTTA GGGTGTGGAA AGTCCCCAGG CTCCCCAGCA GGCAGAAGTA TGCAAAGCAT GCATCTCAAT 3760
3761 TAGTCAGCAA CCAGGTGTGG AAAGTCCCCA GGCTCCCCAG CAGGCAGAAG TATGCAAAGC ATGCATCTCA ATTAGTCAGC 3840
3841 AACCATAGTC CCGCCCCTAA CTCCGCCCAT CCCGCCCCTA ACTCCGCCCA GTTCCGCCCA TTCTCCGCCC CATGGCTGAC 3920
3921 TAATTTTTTT TATTTATGCA GAGGCCGAGG CCGCCTCTGC CTCTGAGCTA TTCCAGAAGT AGTGAGGAGG CTTTTTTGGA 4000
4001 GGCCTAGGCT TTTGCAAAAA GCTCCCGGGA GCTTGTATAT CCATTTTCGG ATCTGATCAA GAGACAGGAT GAGGATCGTT 4080
4081 TCGCATGATT GAACAAGATG GATTGCACGC AGGTTCTCCG GCCGCTTGGG TGGAGAGGCT ATTCGGCTAT GACTGGGCAC 4160
4161 AACAGACAAT CGGCTGCTCT GATGCCGCCG TGTTCCGGCT GTCAGCGCAG GGGCGCCCGG TTCTTTTTGT CAAGACCGAC 4240
4241 CTGTCCGGTG CCCTGAATGA ACTGCAGGAC GAGGCAGCGC GGCTATCGTG GCTGGCCACG ACGGGCGTTC CTTGCGCAGC 4320
4321 TGTGCTCGAC GTTGTCACTG AAGCGGGAAG GGACTGGCTG CTATTGGGCG AAGTGCCGGG GCAGGATCTC CTGTCATCTC 4400
4401 ACCTTGCTCC TGCCGAGAAA GTATCCATCA TGGCTGATGC AATGCGGCGG CTGCATACGC TTGATCCGGC TACCTGCCCA 4480
4481 TTCGACCACC AAGCGAAACA TCGCATCGAG CGAGCACGTA CTCGGATGGA AGCCGGTCTT GTCGATCAGG ATGATCTGGA 4560
4561 CGAAGAGCAT CAGGGGCTCG CGCCAGCCGA ACTGTTCGCC AGGCTCAAGG CGCGCATGCC CGACGGCGAG GATCTCGTCG 4640
4641 TGACCCATGG CGATGCCTGC TTGCCGAATA TCATGGTGGA AAATGGCCGC TTTTCTGGAT TCATCGACTG TGGCCGGCTG 4720
4721 GGTGTGGCGG ACCGCTATCA GGACATAGCG TTGGCTACCC GTGATATTGC TGAAGAGCTT GGCGGCGAAT GGGCTGACCG 4800
4801 CTTCCTCGTG CTTTACGGTA TCGCCGCTCC CGATTCGCAG CGCATCGCCT TCTATCGCCT TCTTGACGAG TTCTTCTGAG 4880
4881 CGGGACTCTG GGGTTCGAAA TGACCGACCA AGCGACGCCC AACCTGCCAT CACGAGATTT CGATTCCACC GCCGCCTTCT 4960
4961 ATGAAAGGTT GGGCTTCGGA ATCGTTTTCC GGGACGCCGG CTGGATGATC CTCCAGCGCG GGGATCTCAT GCTGGAGTTC 5040
5041 TTCGCCCACC CCAACTTGTT TATTGCAGCT TATAATGGTT ACAAATAAAG CAATAGCATC ACAAATTTCA CAAATAAAGC 5120
5121 ATTTTTTTCA CTGCATTCTA GTTGTGGTTT GTCCAAACTC ATCAATGTAT CTTATCATGT CTGTATACCG TCGACCTCTA 5200
5201 GCTAGAGCTT GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA 5280
5281 GCCGGAAGCA TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC 5360
5361 TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGGC 5440
5441 GCTCTTCCGC TTCCTCGCTC ACTGACTCGC TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG 5520
5521 GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG 5600
5601 TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA 5680
5681 GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC 5760
5761 CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGCT GTAGGTATCT 5840
5841 CAGTTCGGTG TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG 5920
5921 GTAACTATCG TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA 6000
6001 GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGAACAG TATTTGGTAT 6080
6081 CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG 6160
6161 GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG 6240
6241 TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA CCTAGATCCT 6320
6321 TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA 6400
6401 GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTCATC CATAGTTGCC TGACTCCCCG TCGTGTAGAT AACTACGATA 6480
```

```
6481 CGGGAGGGCT TACCATCTGG CCCCAGTGCT GCAATGATAC  CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT 6560
6561 AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT  GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC 6640
6641 GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG  CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC 6720
6721 TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCCAAC  GATCAAGGCG AGTTACATGA TCCCCCATGT TGTGCAAAAA 6800
6801 AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT  AAGTTGGCCG CAGTGTTATC ACTCATGGTT ATGGCAGCAC 6880
6881 TGCATAATTC TCTTACTGTC ATGCCATCCG TAAGATGCTT  TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA 6960
6961 TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA  TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT 7040
7041 GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA  AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA 7120
7121 CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC  CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC 7200
7201 GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC  TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA 7280
7281 GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT  TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC 7360
7361 GAAAAGTGCC ACCTGACGTC                                                                7380
          |   10    |   20    |   30    |   40      |   50    |   60    |   70    |   80    |
```

In the DNA constructs of the present invention, the above SARS-CoV proteins may be substituted by homologues or analogues thereof from any viral isolate or strain, or with a sequence that has conservative substitutions such that the protein maintain their immunogenicity and antigenicity when administered in the form of a nucleic acid composition or polypeptide. In view of the information provided above and in the examples, it is within the skill of the art, without undue experimentation, to combine various SARS-CoV proteins or fragments thereof with a CRT sequence, preferably a human CRT sequence, or a functional variant or fragment thereof that enhances immunogenicity, or the sequence of another endoplasmic reticulum chaperone polypeptide that has similar activity to CRT, to generate a composition that is useful, as, *e.g.*, a chimeric nucleic acid immunogen or vaccine to enhance immunity to a linked antigenic peptide or polypeptide.

Table 2 below shows nucleotide base differences among the TW-1, TOR-2, HKU-39849, CUHK-W1, and the Urbani sequences of SARS-CoV

TABLE 1

	VIRAL ISOLATE/STRAIN					
Base position	TW-1	TOR-2	HKU-39849	CUHK-W1	Urbani	Residue change*
						TW1/Urbani
2,601	T	T	C	T	T	Val/Val
3,165	G	A	A	A	A	Ser/Ser
7,746	G	G	T	T	G	Pro/Pro
7,919	C	C	C	C	T	Ala/Val
9,404	T	T	C	C	T	Val/Ala
9,479	T	T	C	C	T	Val/Ala
16,622	C	C	C	C	T	Ala/Ala
17,564	T	T	G	G	T	Asp/Glu
17,846	C	C	T	T	C	Arg/Arg
19,064	A	A	G	G	G	Glu/Glu
21,721	G	G	A	A	G	Gly/Asp
22,222	T	T	C	C	T	Ile/Thr
23,220	T	G	T	T	T	Ser/Ala
24,872	T	T	T	T	C	Leu/Leu
25,298	G	A	G	G	G	Gly/Arg
26867	T	T	T	T	C	Ser/Pro
27,827	T	T	C	C	T	Cys/Arg

* Indicates a base difference resulting in an amino acid change between TW1 and Urbani.

Techniques for the manipulation of nucleic acids, such as, *e.g.*, generating mutations in sequences, subcloning, labeling probes, sequencing, hybridization and the like are well described in the scientific and patent literature. See, *e.g.*, Sambrook, ed., *MOLECULAR CLONING: A LABORATORY MANUAL* (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, Ausubel, ed. John Wiley & Sons, Inc., New York (1997);

LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Tijssen, ed. Elsevier, N.Y. (1993).

Nucleic acids, vectors, capsids, polypeptides, and the like can be analyzed and quantified by any of a number of general means well known to those of skill in the art. These include, *e.g.*, analytical biochemical methods such as NMR, spectrophotometry, radiography, electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), and hyperdiffusion chromatography, various immunological methods, *e.g.* fluid or gel precipitin reactions, immunodiffusion, immuno-electrophoresis, radioimmunoassays (RIAs), enzyme-linked immunosorbent assays (ELISAs), immuno-fluorescent assays, Southern analysis, Northern analysis, dot-blot analysis, gel electrophoresis (*e.g.*, SDS-PAGE), RT-PCR, quantitative PCR, other nucleic acid or target or signal amplification methods, radiolabeling, scintillation counting, and affinity chromatography.

Amplification of Nucleic Acids

Oligonucleotide primers can be used to amplify nucleic acids to generate fusion protein coding sequences used to practice the invention, to monitor levels of vaccine after *in vivo* administration (*e.g.*, levels of a plasmid or virus), to confirm the presence and phenotype of activated CTLs, and the like. The skilled artisan can select and design suitable oligonucleotide amplification primers using known sequences, *e.g.*, SEQ ID NO:1. Amplification methods are also well known in the art, and include, *e.g.*, polymerase chain reaction, PCR (*PCR Protocols, A Guide to Methods and Applications*, ed. Innis, Academic Press, N.Y. (1990) and *PCR Strategies* (1995), ed. Innis, Academic Press, Inc., N.Y., ligase chain reaction (LCR) (Wu (1989) *Genomics 4*:560; Landegren (1988) *Science 241*:1077; Barringer (1990) Gene 89:117); transcription amplification (Kwoh (1989) *Proc. Natl. Acad. Sci. USA 86*:1173); and, self-sustained sequence replication (Guatelli (1990) *Proc. Natl. Acad. Sci. USA 87*:1874); Qβ replicase amplification (Smith (1997) *J. Clin. Microbiol. 35*:1477-1491; Burg (1996) *Mol. Cell. Probes 10*:257-271) and other RNA polymerase mediated techniques (NASBA, Cangene, Mississauga, Ontario; Berger (1987) *Meth. Enzymol. 152*:307-316; U.S. Patent Nos. 4,683,195 and 4,683,202; Sooknanan (1995) Biotechnology 13:563-564).

Cloning and construction of expression cassettes

Expression cassettes, including plasmids, recombinant viruses (*e.g.*, RNA viruses like the replicons described below) and other vectors encoding the fusion proteins described herein are used to express these polypeptides *in vitro* and *in vivo*. Recombinant nucleic acids are expressed

by a variety of conventional techniques (Roberts (1987) Nature 328:731; Schneider (1995) Protein Expr. Purif. 6435:10; Sambrook, *supra* Tijssen, *supra*; Ausubel, *supra*). Plasmids, vectors, *etc.*, can be isolated from natural sources, obtained from such sources as ATCC or GenBank libraries, or prepared by synthetic or recombinant methods.

The nucleic acids used to practice the invention can be stably or transiently expressed in cells such as episomal expression systems. Selection markers can be incorporated to confer a selectable phenotype on transformed cells. For example, selection markers can code for episomal maintenance and replication such that integration into the host genome is not required. For example, the marker may encode antibiotic resistance, *e.g.*, chloramphenicol, kanamycin, G418, bleomycin, hygromycin) to permit selection of those cells transformed with the desired DNA sequences (Blondelet- Rouault (1997) Gene 190:315-317; Aubrecht (1997) J. Pharmacol. Exp. Ther. 281:992-997).

In Vivo Nucleic Acid Administration

Preferred methods of administration are exemplified herein and are well-known in the art. In one embodiment, a nucleic acid encoding a CRT-SARS peptide epitope chimeric polypeptide are cloned into expression cassettes such as plasmids or other vectors, viruses that can transfect or infect cells *in vitro, ex vivo* and/or *in vivo*. A number of delivery approaches are known, including lipid or liposome based gene delivery (Mannino (1988) *BioTechniques 6*:682-691; U.S. Pat No. 5,279,833), replication-defective retroviral vectors with desired exogenous sequence as part of the retroviral genome (Miller (1990) *Mol. Cell. Biol. 10*:4239; Kolberg (1992) *J. NIH Res. 4*:43; Cornetta (1991) *Hum.. Gene Ther. 2*: 215; Zhang (1996) *Cancer Metastasis Rev. 15*:385-401; Anderson, *Science (*1992) *256*: 808-813; Nabel (1993) *TIBTECH 11*: 211-217; Mitani (1993) *TIBTECH 11*: 162-166; Mulligan (1993) *Science 260A*:926-932; Dillon (1993) *TIBTECH 11*: 167-175; Miller (1992) *Nature 357*: 455-460).

Expression cassettes can also be derived from viral genomes. Vectors which may be employed include recombinantly modified enveloped or non-enveloped DNA and RNA viruses, examples of which are baculoviridae, parvoviridae, picornaviridae, herpesviridae, poxviridae, adenoviridae, picornaviridae or alphaviridae. Chimeric vectors may also be employed which exploit advantageous merits of each of the parent vector properties (Feng (1997) Nature Biotechnology 15:866-870). Such viral genomes may be modified by recombinant DNA techniques to include the gene of interest and may be engineered to be replication-deficient, conditionally replicating or replication-competent. Vectors can be derived from adenoviral,

adeno-associated viral or retroviral genomes. Retroviral vectors can include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), simian immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof (Buchscher (1992) *J. Virol. 66:*2731-2739; Johann (1992) *J. Virol. 66*:1635-1640 (1992); Sommerfelt (1990) Virol. 176:58-59; Wilson (1989) J. Virol. 63:2374-2378; Miller (1991) J. Virol. 65:2220-2224. Adeno-associated virus (AAV)-based vectors can transduce cells for the *in vitro* production of nucleic acids and peptides, and be used in *in vivo* and *ex vivo* therapy procedures (Okada (1996) Gene Ther. 3:957-964; West (1987) Virology 160:38-47; Carter (1989) U.S. Patent No. 4,797,368; Carter *et al.* WO 93/24641 (1993); Kotin (1994) Human Gene Therapy 5:793-801; Muzyczka (1994) J. Clin. Invest. 94:1351).

In vivo administration using self-replicating RNA replicons

In addition to the above-described expression vectors and recombinant viruses, self-replicating RNA replicons can also be used to infect cells or tissues or whole organisms with a fusion protein-expressing nucleic acids of the invention. Thus, the invention also incorporates RNA viruses, including alphavirus genome RNAs such as from Sindbis virus, Semliki Forest virus, Venezuelan equine encephalitis virus, and the like, that have been engineered to allow expression of heterologous RNAs and proteins. High levels of expression of heterologous sequences such as the fusion polypeptides of the invention, are achieved when the viral structural genes are replaced by the heterologous coding sequences.

These recombinant RNAs are self-replicating ("replicons") and can be introduced into cells as naked RNA or DNA. However, they require *trans* complementation to be packaged and released from cells as infectious virion particles. The defective helper RNAs contain the *cis*-acting sequences required for replication as well as an RNA promoter which drives expression of open reading frames. In cells co-transfected with both the replicon and defective helper RNAs, viral nonstructural proteins translated from the replicon RNA allow replication and transcription of the defective helper RNA to produce the virion's structural proteins (Bredenbeek (1993) *J. Virol.* 67:6439-6446).

RNA replicon vaccines may be derived from alphavirus vectors, such as Sindbis virus (family *Togaviridae*) (Xiong (1989) *Science 243*:1188-1191), Semliki Forest virus (Ying (1999) *Nat. Med. 5*:823-827) or Venezuelan equine encephalitis virus (Pushko (1997) *Virology 239*:389-401) vectors. These vaccines are self-replicating and self-limiting and may be administered as either RNA or DNA, which is then transcribed into RNA replicons in

transfected cells or *in vivo* (Berglund (1998) *Nat. Biotechnol. 16*:562-565). Self-replicating RNA infects a diverse range of cell types and allows the expression of the antigen of interest at high levels (Huang (1996) *Curr. Opin. Biotechnol. 7*:531-535). Additionally, self-replicating RNA eventually causes lysis of transfected cells because viral replication is toxic to infected host cells (Frolov (1996) *J. Virol.* 70:1182-1190). These vectors therefore do not raise the concern associated with naked DNA vaccines of integration into the host genome. In one embodiment, the self-replicating RNA replicon comprises a Sindbis virus self-replicating RNA vector SINrep5, as described in detail by Bredenbeek, *supra* and Herrmann (1998) *Biochem. Biophys. Res. Commun.* 253:524-531.

Polypeptides

In other embodiments, the invention is directed to an isolated or recombinant polypeptide comprising at least two domains, wherein the first domain comprises a calreticulin (CRT) polypeptide; and, wherein the second domain comprises an MHC class I-binding peptide epitope of a SARS protein that is antigenic such that an immune response directed against such an epitope leads to any type of protective or prophylactic or therapeutic immunity against the virus. As noted above, the terms "polypeptide," "protein," and "peptide," referring to polypeptides including the CRT, fragments of CRT that bind peptides, and MHC class I-binding peptide epitopes, SARS polypeptides, such as the S, E, M and N proteins to practice the invention. These proteins are disclosed in more detail, including amino acid sequence and encoding nucleic acid sequences, above. The composition of the invention also include "analogues," or "conservative variants" and "mimetics" or "peptidomimetics" with structures and activity that substantially correspond to CRT and SARS protein or epitope(s) thereof. Thus, the terms "conservative variant" or "analogue" or "mimetic" also refer to a polypeptide or peptide which has a modified amino acid sequence, such that the change(s) do not substantially alter the polypeptide's (the conservative variant's) structure and/or activity (ability to bind to "antigenic" peptides, to stimulate an immune response). These include conservatively modified variations of an amino acid sequence, *i.e.,* amino acid substitutions, additions or deletions of those residues that are not critical for protein activity, or substitution of amino acids with residues having similar properties (acidic, basic, positively or negatively charged, polar or non-polar, *etc.*) such that the substitutions of even critical amino acids does not substantially alter structure and/or activity. Conservative substitution tables providing functionally similar amino acids are well known in the art. For example, one exemplary guideline to select conservative substitutions

includes (original residue/substitution): Ala/Gly or Ser; Arg/ Lys; Asn/ Gln or His; Asp/Glu; Cys/Ser; Gln/Asn; Gly/Asp; Gly/Ala or Pro; His/Asn or Gln; Ile/Leu or Val; Leu/Ile or Val; Lys/Arg or Gln or Glu; Met/Leu or Tyr or Ile; Phe/Met or Leu or Tyr; Ser/Thr; Thr/Ser; Trp/Tyr; Tyr/Trp or Phe; Val/Ile or Leu.

An alternative exemplary guideline uses the groups shown in the Table below. For a detailed description of protein chemistry and structure, see Schulz, GE *et al.*, *Principles of Protein Structure*, Springer-Verlag, New York, 1978, and Creighton, T.E., *Proteins: Structure and Molecular Properties*, W.H. Freeman & Co., San Francisco, 1983, which are hereby incorporated by reference. The types of substitutions that may be made in the polypeptides of this invention may be based on analysis of the frequencies of amino acid changes between a homologous protein of different species, defined herein as exchanges within one of the following five groups:

1	Small aliphatic, nonpolar or slightly polar residues	Ala, Ser, Thr (Pro, Gly);
2	Polar, negatively charged residues and their amides	Asp, Asn, Glu, Gln;
3	Polar, positively charged residues	His, Arg, Lys;
4	Large aliphatic, nonpolar residues	Met, Leu, Ile, Val (Cys)
5	Large aromatic residues	Phe, Tyr, Trp.

The three amino acid residues in parentheses above have special roles in protein architecture. Gly is the only residue lacking a side chain and thus imparts flexibility to the chain. Pro, because of its unusual geometry, tightly constrains the chain. Cys can participate in disulfide bond formation, which is important in protein folding.

More substantial changes in biochemical, functional (or immunological) properties are made by selecting substitutions that are less conservative, such as between, rather than within, the above five groups. Such changes will differ more significantly in their effect on maintaining (a) the structure of the peptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. Examples of such substitutions are (i) substitution of Gly and/or Pro by another amino acid or deletion or insertion of Gly or Pro; (ii) substitution of a hydrophilic residue, *e.g.*, Ser or Thr, for (or by) a hydrophobic residue, *e.g.*,, Leu, Ile, Phe, Val or Ala; (iii) substitution of a Cys residue for (or by) any other residue; (iv) substitution of a residue having an electropositive side chain, *e.g.*,, Lys, Arg or His, for (or by) a residue having an

electronegative charge, *e.g.*,, Glu or Asp; or (v) substitution of a residue having a bulky side chain, *e.g.*, Phe, for (or by) a residue not having such a side chain, *e.g.*, Gly.

One of skill in the art will appreciate that the above-identified substitutions are not the only possible conservative substitutions. For example, for some purposes, all charged amino acids may be considered conservative substitutions for each other whether they are positive or negative. Individual substitutions, deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence can also be considered to yield "conservatively modified variants."

The terms "mimetic" and "peptidomimetic" refer to a synthetic chemical compound that has the necessary structural and/or functional characteristics of a peptide that permits use in the methods of the invention, such as mimicking CRT in interaction with peptides and MHC class I-proteins). The mimetic can be either entirely composed of synthetic, non-natural analogues of amino acids, or, is a combination of partly natural amino acids and partly non-natural analogues. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not substantially alter the mimetics' structure and/or activity. As with conservative variants, routine experimentation will determine whether a mimetic is within the scope of the invention, that its stereochemical structure and/or function is not substantially altered. Peptide mimetics can contain any combination of "non-natural" structural components, typically from three groups: (a) residue linkage groups other than the natural amide bond ("peptide bond"); (b) non-natural residues in place of naturally occurring amino acids; or (c) residues which induce or stabilize a secondary structure, *e.g.*, a β turn, γ turn, β sheet, or α helix conformation. A polypeptide can be characterized as a mimetic when all or some of its residues are joined by chemical bonds other than peptide bonds. Individual peptidomimetic residues can be joined by peptide bonds, other chemical bonds or coupling means, such as glutaraldehyde, N-hydroxysuccinimide esters, bifunctional maleimides, N,N'-dicyclohexylcarbodiimide (DCC) or N,N'-diisopropylcarbodiimide (DIC). Linking groups that are alternatives to peptide bonds include, ketomethylene (-C(=O)-CH_2- for -C(=O)-NH-), aminomethylene (CH_2-NH), ethylene, olefin (CH=CH), ether (CH_2-O), thioether (CH_2-S), tetrazole (CN_4-), thiazole, retroamide, thioamide, or ester (Spatola (1983) in *Chemistry and Biochemistry of Amino Acids, Peptides and Proteins*, Vol. 7, pp 267-357, *Peptide Backbone Modifications*, Marcell Dekker, NY).

The structure of the polypeptides, peptides, other functional derivatives, including mimetics of the present invention are preferably based on structure and amino acid sequence of CRT, preferably human CRT (SEQ ID NO:2, disclosed above) or a SARS-CoV protein such as S, E, M or N as disclosed herein for two viral isolates.

Individual synthetic residues and polypeptides incorporating mimetics can be synthesized using a variety of procedures and methodologies well known in the art, *e.g.*, *Organic Syntheses Collective Volumes*, Gilman *et al.* (eds) John Wiley & Sons, Inc., NY. Polypeptides incorporating mimetics can also be made using solid phase synthetic procedures (*e.g.*, U.S. Pat. No. 5,422,426). Peptides and peptide mimetics of the invention can also be synthesized using combinatorial methodologies. Various techniques for generation of peptide and peptidomimetic libraries are well known *e.g.*, multipin, tea bag, and split-couple-mix techniques (al-Obeidi (1998) Mol. Biotechnol. 9:205-223; Hruby (1997) Curr. Opin. Chem. Biol. 1:114-119; Ostergaard (1997) Mol. Divers. 3:17-27; Ostresh (1996) Methods Enzymol. 267:220-234). Modified polypeptide and peptides can be further produced by chemical modification (Belousov (1997) Nucleic Acids Res. 25:3440-3444; Frenkel (1995) Free Radic. Biol. Med. 19:373-380; Blommers (1994) Biochemistry 33:7886-7896).

The peptides can also be synthesized, whole or in part, using conventional chemical synthesis (Caruthers (1980) *Nucleic Acids Res. Symp. Ser.* 215-223; Horn (1980) *Nucleic Acids Res. Symp. Ser.* 225-232; Banga, A.K., *Therapeutic Peptides and Proteins, Formulation, Processing and Delivery Systems* (1995) Technomic Publishing Co., Lancaster, PA. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge (1995) Science 269:202; Merrifield (1997) Methods Enzymol. 289:3-13) and automated synthesis, *e.g.*, using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the manufacturer' instructions.

In one embodiment of the invention, peptide-binding fragments or "sub-sequences" of CRT are used. In another embodiment, other peptides that bind to MHC proteins, preferably MHC Class I proteins, are used. Such peptides can be derived from any polypeptide, particularly, from a known pathogen, or it can be entirely synthetic). Methods for determining whether, and to what extent, a peptide binds to a CRT or a CRT fragment, or an MHC protein are routine in the art (Jensen (1999) *Immunol. Rev.* 172:229-238; Zhang (1998) *J. Mol. Biol.* 281:929-947; Morgan (1997) *Protein Sci* 6:1771-1773; Fugger (1996) *Mol. Med.* 2:181-188;

Sette (1994) *Mol. Immunol.* 31:813-822; Elvin (1993) *J. Immunol. Meth.* 158:161-171; U.S. Patent Nos. 6,048,530; 6,037,135; 6,033,669; 6,007,820).

Formulation and Administration of Pharmaceutical or Immunological Compositions

In various embodiments of the invention, polypeptides, nucleic acids, expression cassettes, cells, and particles, are administered to an individual as pharmacological compositions in amounts sufficient to induce an antigen-specific immune response (*e.g.*, a CTL response, see Examples, below) in the individual.

Pharmaceutically acceptable carriers and formulations for nucleic acids, peptides and polypeptides are known to the skilled artisan and are described in detail in the scientific and patent literature, see *e.g.*, the latest edition of Remington's Pharmaceutical Science, Mack Publishing Company, Easton, PA ("Remington's"); Banga; Putney (1998) Nat. Biotechnol. 16:153-157; Patton (1998) Biotechniques 16:141-143; Edwards (1997) Science 276: 1868-1871; U.S. Patent Nos. 5,780,431; 5,770,700; 5,770,201.

The nucleic acids and polypeptides used in the methods of the invention can be delivered alone or as pharmaceutical compositions by any means known in the art, *e.g.*, systemically, regionally, or locally; by intraarterial, intrathecal (IT), intravenous (IV), parenteral, intra-pleural cavity, topical, oral, or local administration, as subcutaneous, intra-tracheal (*e.g.*, by aerosol) or transmucosal (*e.g.*, buccal, bladder, vaginal, uterine, rectal, nasal mucosa). Actual methods for delivering compositions will be known or apparent to those skilled in the art and are described in detail in the scientific and patent literature, see *e.g.*, Remington's.

The pharmaceutical compositions can be administered by any protocol and in a variety of unit dosage forms depending upon the method and route and frequency of administration, whether other drugs are being administered, the individual's response, and the like. Dosages for typical nucleic acid, peptide and polypeptide pharmaceutical compositions are well known to those of skill in the art. Such dosages may be adjusted depending on a variety of factors, *e.g.*, the initial responses (*e.g.*, number and activity of CTLs induced, tumor shrinkage, anti-viral activity measured as lysis of virus-infected cells or reduction of virus titer, and the like), the particular therapeutic context, patient health and tolerance. The amount of pharmaceutical composition adequate to induce the desired response is defined as a "therapeutically effective dose." The dosage schedule and amounts effective for this use, *i.e.*, the "dosing regimen," will depend upon a variety of factors, including, *e.g.*, the diseases or conditions to be treated or

prevented by the immunization, the general state of the patient's health, the patient's physical status, age, pharmaceutical formulation and concentration of pharmaceutical composition, and the like. The dosage regimen also takes into consideration pharmacokinetics, *i.e.*, the pharmaceutical composition's rate of absorption, bioavailability, metabolism, clearance, and the like (Remington). Dosages can be determined empirically, *e.g.*, by assessing the abatement or amelioration of symptoms, or, by objective criteria, *e.g.*, measuring levels of antigen-specific CTLs. As noted above, a single or multiple administrations can be administered depending on the dosage and frequency as required and tolerated by the patient. The pharmaceutical compositions can be administered alone or in conjunction with other therapeutic treatments, or, as prophylactic immunization.

Ex vivo treatment and re-administration of APCs

In various embodiments of the invention, the nucleic acids and polypeptides of the invention are introduced into the individual by *ex vivo* treatment of antigen presenting cells (APCs), followed by administration of the manipulated APCs. In one embodiment, APCs are transduced (transfected) or infected with fusion protein-encoding nucleic acids of the invention; afterwards, the APCs are administered to the individual. In another embodiment, the APCs are stimulated with fusion proteins of the invention (purified or as a cell lysate from cells transfected and expressing a recombinant fusion protein *in vivo*). Afterward this "pulsing, the APCs are administered to the individual.

The fusion proteins can be in any form, *e.g.*, as purified or synthetic polypeptides, as crude cell lysates (from transfected cells making recombinant fusion protein), and the like. The APC can be an MHC-matched cell (a tissue-typed cell). The APC can be a tissue-cultured cell or it can be an APC isolated from the individual to be treated and re-administered after *ex vivo* stimulation. Any APC can be used, as described above. Methods of isolating APCs, *ex vivo* treatment in culture, and re-administration are well known in the art (U.S. Patent Nos. 5,192,537; 5,665,350; 5,728,388; 5,888,705; 5,962,320; 6,017,527; 6,027,488).

Kits

The invention provides kits that contain the pharmaceutical or immunogenic compositions of the invention, as described above, to practice the methods of the invention. In alternative embodiments, the kits can contain recombinant or synthetic chimeric polypeptides comprising a first domain comprising an ER chaperone polypeptide and a second domain comprising an antigenic peptide of the SARS CoV, *e.g.*, a CRT-Class I-binding peptide epitope

fusion protein; or, the nucleic acids encoding them, *e.g.*, in the form of naked DNA (*e.g.*, plasmids), viruses (*e.g.* alphavirus-derived "replicons" including Sindbis virus replicons) and the like. The kit can contain instructional material teaching methodologies, *e.g.*, means to administer the compositions used to practice the invention, means to inject or infect cells or patients or animals with the nucleic acids or polypeptides of the invention, means to monitor the resultant immune response and assess the reaction of the individual to which the compositions have been administered, and the like.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims.

EXAMPLES

The following examples are offered to illustrate, but not to limit the claimed invention.

EXAMPLE 1

DNA Vaccines Targeting the Nucleocapsid Protein of SARS-CoV

This Example is built upon the prior discovery of the present inventors that DNA vaccination with antigen linked to calreticulin (CRT) dramatically enhances MHC class I presentation of a linked antigen to $CD8^+$ T cells. In this study, they employed a CRT-based enhancement strategy to create effective DNA vaccines using SARS-CoV nucleocapsid (N) protein as a target antigen. Vaccination with naked CRT/N DNA generated the most potent N-specific humoral and T cell-mediated immune responses in vaccinated C57BL/6 mice among all of the DNA constructs compared here. Animals vaccinated with CRT/N DNA were capable of significantly reducing the titer of challenging vaccinia expressing the N protein of the SARS virus. These results show that a DNA composition encoding CRT linked to a SARS-CoV antigen N can generate strong **N**-specific humoral and cellular immunity that can control infection with SARS-CoV.

Materials and Methods

Plasmid DNA Constructs and DNA Preparation

The current study employed the mammalian expression vector, pcDNA3.1/myc-His (-) (Invitrogen, Carlsbad, CA). For the generation of pcDNA3-N-myc, the DNA fragment encoding SARS-Co V nucleocapsid was amplified with PCR using a set of primers:

5'-AAAGAATTCATGTCTGATAATGGACCCCAATC-3', SEQ ID NO:97

5'-TTTGGTACCTGCCTGAGTTGAATCAGCAGA-3' SEQ ID NO:98

and pGEX-1-NC-G3 (Huang, LR *et al.*, 2004, J Med Virol. 73:338-346) as a template. The amplified product was further cloned into the EcoRI/KpnI sites of pcDNA3.1/*myc*-His (-) vector. To generate pcDNA3-CRT-myc, CRT DNA segment was isolated from pcDNA3-CRT (Cheng, W.-F. *et al.*, 2001, *J. Clinical Invest. 108*:669-678) and cloned into the XhoI/EcoRI sites of pcDNA3.1/*myc*-His (-). For the generation of pcDNA3-CRT/N-myc, the amplified N DNA was cloned into the EcoRI/KpnI sites of pcDNA3-CRT-myc. The accuracy of these constructs was confirmed by DNA sequencing. The DNA was amplified in E. coli DH5α and purified as described previously (Chen, C.-H. *et al.*, 2000, Cancer Research 60:1035-1042; Wu *et al.*, PCT Publication WO 01/29233).

Generation of Bacteria-Derived SARS-CoV N Protein

cDNA encoding SARS nucleocapsid protein was generated by reverse transcription of SARS coronavirus TW1 (18) (Hsueh, PR, 2003, *Emerg Infect Dis 9*:1163-1167;) (accession no. YA291451) using Superscript II (Invitrogen, Carlsbad, CA) followed by amplification using platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA) as described previously (Huang *et al., supra*). The oligonucleotide primers for SARS-CoV N protein were

5'-ATGTCTGATAATGGACCCCA-3' (forward, nt28120-nt28139) SEQ ID NO:99; and

5'-TTATGCCTGAGTTGAATCAG-3' (reversed, nt29369-nt29388). SEQ ID NO:100

The DNA fragment encoding N protein was cloned into pGEX-1plasmid (Amersham Pharmacia Biotech, Little Chalfont, England) to generate pGEX-1-NC-G3 (Huang *et al., supra*) for recombinant protein expression. *E. coli* BL-21 were transformed with pGEX-1 or pGEX-1-NC-G3 plasmids and grown overnight in LB medium containing 50μg/ml ampicillin to the midlog phase. Cells transformed with GST or GST-N fusion constructs were directly induced with 0.25 mM IPTG (isopropyl-β-D-thiogalactoside) for 3 hours at 30 °C. Cells were collected by centrifugation and then resuspended in TNE buffer (50mM Tris, pH 8.0, 0.15M NaCl, 1mM EDTA, and 1mM PMSF), about 1ml per 25OD_{600} cells. The fusion protein solubility was

determined by sonication, and centrifugation followed by SDS–PAGE separation of both the supernatant and pellet fractions. In larger volume of culture (~3 liters), cells were lysed by microfluidizer. Lysates prepared from the large batch were incubated with TNE equilibrated glutathione resin. Bound protein was eluted by 10mM reduced glutathione in 50mM Tris (pH 8.0) buffer. The eluted and purified fractions were used for Western blot analysis and as the coating antigen for ELISA assay.

Western Blot Analysis

The expression of N protein in 293 cells transfected with pcDNA3.1/*myc*-His (-) encoding no insert, CRT, N, or CRT/N DNA was characterized by western blot analysis. 20 μg of DNA were transfected into 5x10^6 293 cells using lipofectamine 2000 (Life Technologies, Rockville, MD). 24 hr after transfection, cells were lysed with protein extraction reagent (Pierce, Rockford, IL). Equal amounts of proteins (50 μg) were loaded and separated by SDS-PAGE using a 10% polyacrylamide gel. For the characterization of bacteria-derived N protein, 1 μg of purified GST-N fusion protein was loaded and separated by SDS-PAGE using a 10% polyacrylamide gel. The gels were electroblotted to a polyvinylidene difluoride membrane (Bio-Rad, Hercules, CA). Blots were blocked with PBS/0.05% Tween 20 (TTBS) containing 5% nonfat milk for 2 hr at room temperature. Membranes were probed with rabbit anti-GST-N sera (Huang *et al.*, *supra*) at 1:1000 dilution in TTBS for 2 hr, washed four times with TTBS, and then incubated with goat anti-rabbit IgG conjugated to horseradish peroxidase (Zymed, San Francisco, CA) at 1:1000 dilution in TTBS containing 5% nonfat milk. Membranes were washed four times with TTBS and developed using enhanced Hyperfilm-enhanced chemiluminescence (Amersham, Piscataway, NJ).

Mice

Six- to eight-week-old female C57BL/6 mice were purchased from the National Cancer Institute (Frederick, Maryland) and kept in the oncology animal facility of the Johns Hopkins Hospital (Baltimore, Maryland). All animal procedures were performed according to approved protocols and in accordance with recommendations for the proper use and care of laboratory animals.

DNA Vaccination

DNA-coated gold particles were prepared according to a previously described protocol (Chen *et al., supra*). DNA-coated gold particles were delivered to the shaved abdominal region of mice using a helium-driven gene gun (BioRad, Hercules, CA) with a discharge pressure of 400 p.s.i. C57BL/6 mice were immunized with 2 µg of the plasmid encoding no insert, CRT, N, or CRT/N protein. The mice received two boosters with the same dose at a one week interval.

Enzyme-Linked Immunoabsorbent Assay (ELISA)

The presence of SARS-CoV N-specific antibodies in the sera from CRT/N DNA-vaccinated C57BL/6 mice (5 per group) were determined by ELISA using microwell plates coated with bacteria-derived recombinant GST-N protein. Purified GST-N protein was diluted to 1µg/ml with o.o5 M carbonate buffer (pH 9.6), and 0.1 ml/well was added to 96-well microtiter plates. Purified GST protein was used as negative control. The plates were incubated overnight at 4 °C, washed with phosphate buffered saline (PBS) – 0.05% Tween 20 (PT), incubated with (0.1 ml/well) PT-2% bovine serum albumin (PBT) for 60 minutes at 37°C and washed again with PT. Serial dilutions of the tested sera were added (0.1 ml/well) and the plates were incubated for 60 minutes at 37°C. The plates were washed with PT and were incubated with (0.1 ml/well) alkaline phosphatase-conjugated rabbit anti-mouse antibodies (Zymed, San Francisco, CA) for 30 minutes at 37°C. The plates were washed with PT and incubated with (0.1 ml/well) alkaline phosphatase substrate (according to Sigma instructions) for 60 minutes at 37 °C. Plates were read on a MicroElisa reader at a wavelength of 450 nm. Reading higher than 3-fold negative controls were scored as positive reactions.

Intracellular Cytokine Staining and Flow Cytometry Analysis

In order to assess the ability of our DNA vaccine encoding SARS-CoV N protein to elicit an N-specific CD8+ T cell response, we sought to identify the MHC class I-restricted CTL epitope of the SARS-CoV N protein. Using the BioInformatics & Molecular Analysis Section (BIMAS) for D^b and K^b peptide binding predictions (URL is bimas.cit.nih.gov/molbio/hla_bind/) and the SYFPEITHI database of MHC ligands and peptide motifs (URL is syfpeithi.bmi-heidelberg.com/), we analyzed various peptides of eight, nine, or ten residues and determined their sequences, positions, and scores, and eventually generated 7 potential peptides for our studies (see Table 3). We used splenocytes from C57BL/6 mice vaccinated with CRT/N DNA

for the characterization of these candidate peptides. Splenocytes were harvested from mice one week after the last vaccination. Prior to intracellular cytokine staining, 4×10^6 pooled splenocytes from the vaccinated mice were incubated for 16 hours with 1 μg/ml of each candidate peptide for detecting N-specific $CD8^+$ T cell precursors. Intracellular IFN-γ staining and flow cytometry analysis were performed as described previously. Flow cytometry analysis was performed on a Becton-Dickinson FACScan with CELLQuest software (Becton Dickinson Immunocytometry System, Mountain View, CA).

To characterize the various DNA vaccines in eliciting an N-specific CD8+ T cell response, splenocytes from the various vaccinated mice (5 per group) were incubated with 1 μg/ml of N peptide (aa 346-354, QFKDNVILL; SEQ ID NO:31) for 16 hours. Intracellular IFN-γ staining and flow cytometry analysis were performed as described above.

Generation and Characterization of Recombinant Vaccinia

The recombinant vaccinia virus was generated using a protocol similar to that described previously Wu, T.-C., *et al.,* 1995, *Proc. Natl. Acad. Sci. 92*:11671-11675). Briefly, the DNA fragment encoding SARS-Co V nucleocapsid was amplified with PCR using a set of primers:

5'-AAAGCATGCATGTCTGATAATGGACCCCAATC-3' (SEQ ID NO:32)
5'-TTTGGTACCTTATGCCTGAGTTGAATCAGCAGA-3' (SEQ ID NO:32)and
pGEX-1-NC-G3 as a template. The amplified product was further cloned into sphI/KpnI sites of pSCIIMCS2. This construct was transfected into Vac-WT infected CV-1 using Lipofectamine 2000. The recombinant vaccinia viruses were isolated as in Wu *et al., supra*. Plaque-purified recombinant vaccinia viruses were checked for the expression of N protein by flow cytometry analysis, immunofluorescence staining, and Western blot analysis using rabbit anti-GST-N sera (Huang *et al., supra*). For the detection of the expression of SARS-CoV N protein in TK^- cells infected with Vac-N by flow cytometry analysis, the vaccinia-infected cells were incubated with rabbit anti-GST-N sera at 1:100 dilution in 1x Perm (PharMingen, San Diego, CA) for 30 min after fixation with Cytofix/Cytoperm (PharMingen, San Diego, CA), washed four times with 1X PBS, and then incubated with FITC-labeled goat anti-rabbit IgG (Jackson ImmunoReseach Laboratories, West Grove, PA) at 1:1000 dilution. Western blot analysis was performed as described above.

The Vac-WT and Vac-N were amplified by infecting TK$^-$ cells *in vitro* according to a standard protocol. Titer was determined by plaque assay using BSC-1 cells. The viral stocks were preserved at –70°C prior to vaccination. Before use, the virus was thawed, trypsinized with 1/10 volume of trypsin/EDTA in 37°C water bath for 30 min, and diluted with minimal essential medium (MEM) to the final concentration of 1×10^8 plaque-forming units (PFU)/ml.

Immunofluorescence Staining for N Protein Expression

Immunofluorescence staining was performed using a protocol similar to what has been described previously (Cheng, WF *et al.*, 2002, *Hum Gene Ther 13*:553-568). Briefly, Tk$^-$ cells were cultured in 8-well culture chamber slides (Nalge Nunc Int., Naperville, IL) until they reached 50% confluence. The cells were infected with Vac-N or Vac WT at 10 m.o.i. to evaluate the expression of N protein. After 24 hours of infection, cells were fixed and permeabilized with Cytofix/Cytoperm (Pharmingen) for 30 min. Rabbit anti-N sera was added into the chamber at a dilution of 1:100 and incubated for 30 min. Diluted FITC goat anti-rabbit IgG (10 μg/ml, Jackson ImmunoReseach Laboratories, West Grove, PA), was added and incubated for 30 min. The slides were mounted and observed immediately under a fluorescence microscope.

***In Vivo* Challenge with Recombinant Vaccinia Virus**

For the local challenge experiment, the immunized mice were anesthetized and infected with 2×10^6 PFU/mouse of Vac-WT or Vac-N in 20 μl by intranasal instillation 1 week after the final immunization. For the systemic challenge experiment, the immunized mice were infected with 1×10^7PFU/mouse of Vac-N in 100 μl by intravenous injection 1 week after the final immunization. Five mice were used for each vaccinated group. To determine virus titers in lungs, mice were sacrificed 5 days after challenge. Both lungs were harvested, homogenized in 1 ml of MEM containing 2.5% fetal bovine serum, and subjected to three rounds of freezing and thawing before the titer of virus was determined by plaque assay.

Statistical Analysis

All data expressed as means ± SEM are from one experiment of at least two experiments performed. Data for intracellular cytokine staining with flow cytometry analysis and *in vivo* viral challenge experiments were evaluated by analysis of variance (ANOVA). Comparisons between individual data points were made using a student's t-test.

Results

Characterization of N protein in cells transfected with the various DNA vaccines.

In order to characterize the expression of the SARS-CoV N protein in 293 cells transfected with the various DNA constructs, we performed a Western blot analysis, using cell lysates derived from DNA-transfected cells. Rabbit anti-GST-N sera were used for Western blot analysis. As shown in **Figure 1**, lysate from 293 cells transfected with N DNA revealed a protein band with a size of approximately M_r 48,000 corresponding to N protein in Lane 3. Lysate from 293 cells transfected with CRT/N DNA revealed a protein band with a size of approximately M_r 90,000 corresponding to the chimeric CRT/N protein in Lane 4. In contrast, N protein was not detected in lysates from 293 cells transfected with plasmid DNA with no insert (lane 1) or CRT DNA (lane 2). Our data indicated that N DNA-transfected cells exhibited levels of N protein expression comparable to CRT/N DNA-transfected cells.

Vaccination with CRT/N DNA significantly enhances N-specific antibody responses.

To evaluate the humoral immune response to DNA vaccines encoding SARS-CoV N protein, we performed ELISA analysis using bacteria-derived GST-N fusion protein and sera from mice vaccinated with the various DNA vaccines. As shown in **Figures 2A** and **2B**, recombinant GST-N protein was purified from bacteria. The purification of bacteria-derived GST-N protein was demonstrated by gel electrophoresis (**Figure 2A**). The confirmation of GST-N protein was demonstrated by Western blot analysis with rabbit anti-GST-N sera (**Figure 2B**). We used the bacteria-derived GST-N protein for our ELISA. As shown in **Figure 2C**, mice vaccinated with CRT/N DNA generated the highest titer of N-specific antibody responses among mice vaccinated with the various DNA vaccines. Furthermore, ELISA to determine the subtype of IgG antibody showed significantly higher titer of N-specific IgG1 Ab than N-specific IgG2a in serum from mice vaccinated with N or CRT/N DNA (**Figure 2D**). We also used purified GST protein as a control for our ELISA. Sera from vaccinated mice only generated background level of color changes against GST (data not shown). These data show that vaccination with CRT/N DNA elicits a significantly stronger N-specific humoral immune response than vaccination with N DNA. This suggests that the linkage of CRT to N protein in a DNA vaccine enhances N-specific antibody production in vaccinated mice.

Vaccination with CRT/N DNA significantly improved SARS-CoV N-specific CD8+ T cell-mediated immune responses.

T cell mediated immunity has been shown to be important for control of viral infection. In order to develop quantitative assays for characterizing N-specific CD8+ T cell mediated immune responses, we sought to identify the MHC class I-restricted CTL epitope of the SARS-CoV N protein. Using the BioInformatics & Molecular Analysis Section (BIMAS) for D^b and K^b peptide binding predictions (http://bimas.cit.nih.gov/molbio/hla_bind/) and the SYFPEITHI database of MHC ligands and peptide motifs (http://syfpeithi.bmi-heidelberg.com/), we identified several potential candidate peptides for SARS-CoV N protein in C57BL/6 mice. **Table 3** shows their sequences, positions, and scores.

Table 3.Candidate CTL epitopes for SARS coronavirus nucleocapsid protein

Peptide name	MHC Class I	length	Peptide position	Peptide sequence	SEQ ID NO:	BIMAS score	SYFPEITHI score
$N_{346-354}$	H-2D^b	9	346-354	QFKDNVILL	31	60	20
$N_{351-359}$	H-2D^b	9	351-359	VILLNKHID	34	33	11
$N_{352-360}$	H-2D^b	9	352-360	ILLNKHIDA	35	n/a	2
$N_{202-211}$	H-2D^b	10	202-211	SSRGNSPARM	36	n/a	24
$N_{122-131}$	H-2D^b	10	122-131	LPYGANKEGI	37	200	n/a
N_{50-57}	H-2K^b	8	50-57	TASWFTAL	38	11	22
$N_{311-318}$	H-2K^b	8	311-318	SASAFFGM	39	11	18

We then synthesized these peptides and characterized their ability to activate N-specific CD8+ T cells using splenocytes harvested from mice vaccinated with the various DNA vaccines. As shown in **Figure 3A,** using intracellular cytokine staining followed by flow cytometry analysis, we showed that a D^b-restricted 9mer peptide positioned at aa 346-354 (QFKDNVILL; SEQ ID NO:31) of N protein was able to activate significantly more N-specific CD8+ T cells in splenocytes from mice vaccinated with CRT/N DNA than the other epitopes ($p<0.05$). In comparison, the N peptide (aa 351-359, VILLNKHID; SEQ ID NO:34) only activated N-specific CD8+ T cells in splenocytes from mice vaccinated with CRT/N DNA to a slightly higher level than the background level. The other five peptides were not able to activate N-specific CD8+ T cells in splenocytes from mice vaccinated with CRT/N DNA **(Figure 3A)**. Thus, the N peptide (aa 346-354, QFKDNVILL; SEQ ID NO:31) likely represents an H-2 D^b-restricted CTL epitope for SARS-CoV N protein. Our results also showed that mice vaccinated with CRT/N DNA generated significantly more N-specific $CD8^+$ T cells than mice vaccinated with N DNA **(Figure 3B)** ($p<0.05$). Thus, our data suggest that the linkage of CRT to N protein in a DNA vaccine enhances N-specific CD8+ T cell mediated immune responses in vaccinated mice.

Recombinant vaccinia expressing SARS-CoV N protein as surrogate virus for vaccine studies

Certain factors preclude the usage of live SARS-CoV for our vaccine efficacy studies. Thus, we generated vaccinia virus expressing SARS-CoV N protein as a surrogate virus for our vaccine efficacy studies. To demonstrate the expression of SARS-CoV N protein expression, we infected 293 cells with vaccinia virus encoding N (Vac-N) and confirmed N expression via flow cytometry analysis, immunofluorescence staining, and Western blot analysis using rabbit anti-GST-N sera (**Figure 4**). 293 cells infected with wild-type vaccinia (Vac-WT) were used as a negative control. All three assays determined that 293 cells infected with Vac-N expressed significant levels of N protein and that 293 cells infected with Vac-WT did not express N protein.

Vaccination with CRT/N DNA results in the greatest reduction of titer of recombinant vaccinia virus expressing N protein.

The ability of a vaccine to successfully protect against viral challenge is an essential measure of its efficacy. To test the ability of our DNA vaccines encoding SARS-COV N protein to protect against viral challenge, we vaccinated mice with DNA encoding CRT/N, N, CRT or no insert and challenged these mice with Vac-N or Vac-WT **intranasally or intravenously** one week after the last vaccination. As shown in **Figure 5A**, while no difference in Vac-WT titer was observed among mice vaccinated with any of the DNA vaccines, we found significantly lower titers of Vac-N in lungs of mice vaccinated with DNA encoding N than in lungs of mice vaccinated with DNA encoding CRT, or no insert (intranasal: $p<0.009$; intravenous: $p<0.033$). More importantly, mice vaccinated with DNA encoding CRT/N exhibited a significantly reduced titer of Vac-N in their lungs when compared to mice vaccinated with DNA encoding N (intranasal: $p<0.013$; intravenous: $p<0.006$). These data indicate that vaccination with CRT/N DNA can reduce titer of vaccinia expressing SARS-CoV N protein to a greater degree than vaccination with N DNA. Thus, vaccination with CRT/N DNA may generate the best protection against intranasal or intravenous challenge with viruses expressing SARS-CoV N protein.

Discussion

Vaccination with CRT/N DNA can elicit SARS-CoV nucleocapsid-specific humoral and cellular immune responses, and our results suggest that these responses can significantly reduce the titer of challenging vaccinia virus expressing N protein. These results also indicate that the linkage of CRT DNA to N DNA leads to enhanced DNA vaccine potency against a virus expressing a SARS-CoV protein. This is consistent with our previous studies using a different

antigen (HPV-16 E7). Thus, the ability of the CRT strategy to enhance cellular and humoral immune responses has been confirmed in two distinct antigenic systems. This indicates that a similar DNA vaccine strategy may prove effective against other antigenic proteins of SARS-CoV, such as the S, E, or M proteins.

The observed enhancement of the humoral immune response against the N protein of SARS-CoV in mice vaccinated with the chimeric CRT/N DNA vaccine may not be useful for SARS-CoV neutralization given the location of the N protein inside the viral envelope. Thus, N-specific antibodies may not be able to cross the envelope to bind with the nucleocapsid protein to abolish the infection. In comparison, SARS-CoV S, E, and M proteins are expressed on the envelope surface, and neutralizing antibodies against these proteins may thus be able to neutralize SARS-CoV infection. This raises the possibility that a DNA vaccine strategy employing CRT linked to the S, E, or M proteins may elicit effective neutralizing antibodies as well as potent T cell responses against infection by live SARS-CoV (see following Examples).

While the humoral immune response may represent an effective means of generating protection from SARS-CoV infection, it may also lead to an antibody-dependent enhancement (ADE) reaction. In ADE, virus-specific antibodies have been shown to interact with the Fc and/or complement receptors to enhance viral entry into host immune cells, such as granulocytic cells and monocytes/macrophages. The ADE phenomenon has been observed in at least one coronaviral system. It should therefore be considered when designing a vaccine against SARS-CoV. If the ADE phenomenon is observed in SARS-CoV infection or vaccination, N protein may be the logical choice for a target antigen, as antibodies against N will be unlikely to lead to ADE. This is due to the fact that the N protein is not expressed on the viral envelope and thus antibodies against N will probably not be able to facilitate viral entry.

We observed significant enhancement of the N-specific CD8+ T cell response as a result of linkage of N protein to CRT in a DNA vaccine. The percentage of N-specific CD8+ T cells in CRT/N DNA-vaccinated mice may potentially be further improved by coadministration with DNA encoding an antiapoptotic protein. Coadministration of DNA encoding BCL-xL with DNA encoding E7/HSP70, CRT/E7, or Sig/E7/LAMP-1 resulted in further enhancement of the E7-specific CD8+ T cell response for all three constructs. Because intracellular targeting and anti-apoptotic strategies modify DCs via different mechanisms, it is potentially feasible to combine anti-apoptotic strategies for prolonging DC life with CRT for enhancing MHC class I

processing and presentation of SARS-CoV antigen by DCs to further enhance DNA vaccine potency.

In this study we used vaccinia virus expressing N protein of SARS-CoV as a surrogate virus for assaying the vaccine efficacy in our study because SARS-CoV, having mainly been isolated in Asia, is difficult to obtain in the United States. More importantly, the handling of live SARS-CoV is potentially extremely hazardous, whereas the handling of recombinant vaccinia is relatively safe. For these reasons, we generated vaccinia expressing SARS-CoV N protein for use as a surrogate viral challenge model. The development of such a model for testing of our vaccine strategy is not without precedent, as vaccinia virus has been previously used in several prior studies as a substitute viral challenge model. While these studies may show a good correlation between the reduction of vaccinia titer and vaccine potency, it would preferable for our research to explore vaccine efficacy against live SARS-CoV virus in a near-human model. A potential animal model is *Macaca Fascicularis*, which has been shown to be susceptible to live SARS-CoV infection and demonstrate pulmonary pathology similar to humans.

DNA vaccination can successfully elicit SARS-CoV N-specific humoral and CD8+ T cell responses in vaccinated mice, and vaccination with CRT/N DNA can significantly enhance both humoral and cellular immune responses when compared to vaccination with N DNA. These enhanced immune responses resulting from linkage of antigen to CRT correlate with a strong reduction of titer of challenging vaccinia expressing N protein in mice vaccinated with CRT/N DNA. While N protein may not be able to elicit an effective neutralizing antibody response against live SARS-CoV, we have shown that it is capable of eliciting a SARS-CoV antigen-specific CD8+ T cell response that results in a significant reduction of titer of challenging vaccinia when linked to CRT in a DNA vaccine. This makes the present CRT/N DNA vaccine a potential candidate for future clinical translation. Furthermore, the CRT DNA vaccination strategy is applicable to envelope-associated SARS-CoV proteins, such as S, E, or M proteins, for elicitation of both neutralizing antibodies against SARS-CoV and SARS-CoV antigen-specific CTLs.

EXAMPLE 2

DNA Vaccines Targeting the Spike Protein (S) of SARS -CoV

Materials and Methods

Plasmid DNA Constructs and DNA Preparation

For the generation of pRSETA-S, the DNA fragment encoding the full-length S protein of SARS-CoV was amplified using a set of primers

5'- cggatccatgtttattttcttattatttct -3' (SEQ ID NO:40) and

5'- cagaattcttatgtgtaatgtaatttgaca -3' (SEQ ID NO:41)

and cDNA from TW-1 strain of SARS- CoV. The amplified product was cloned into the BamHI/EcoRI of pRSETA (Invitrogen, Carslbad, CA).

For the generation of pcDNA3-S, a DNA fragment encoding S was isolated from pRSETA-S and further cloned into the BamHI/EcoRI sites of pcDNA3.1(+) vector (Invitrogen, Carlsbad, CA).

For the generation of pcDNA3 encoding SARS-CoV S1, Si or S2, the DNA fragments encoding S1, Si or S2 DNA fragments were amplified with PCR using the following set of primers:

S1 5'-ccggatccatgtttattttcttattat-3', (SEQ ID NO:42)
5'-ccgaattcttaagacatagtataagccacaatag-3'), (SEQ ID NO:43)

Si 5'-cttggatccatgggttgtgtccttgcttgg-3', (SEQ ID NO:44)
5'-ccgaattcttacatgaagccagcatcagcgag) and (SEQ ID NO:45)

S2 5'-ccggatccatgatgttaggtgctgatagttcaattg-3', (SEQ ID NO:46)
5'-gccgaattcttatgtgtaatgtaatttg-3'), (SEQ ID NO:47)

and pRSETA-S as a template. The amplified products were further cloned into the BamHI/EcoRI sites of pcDNA3.1 (+) vector.

pcDNA3-CRT has been described previously (Cheng, 2001, *supra)*. For the generation of pcDNA3-CRT/S1, the CRT DNA fragment was amplified with PCR using a set of primers:

5'- ggtcttaagatgctgctccctgtgccgctg - 3', (SEQ ID NO:48)

5'- caaagatctcagctcgtccttggcctggc - 3' (SEQ ID NO:49)

and pcDNA3-CRT as a template. The amplified CRT was cloned into the AflII/BamH I sites of pcDNA3-S1. For the generation of pMSCV-S, a DNA fragment encoding S was isolated from pRSETA-S and further cloned into the BglII/EcoRI sites of pMSCV vector (Invitrogen, Carlsbad, CA). The accuracy of these constructs was confirmed by DNA sequencing. The DNA was amplified in E. coli DH5α and purified as described previously.

Cell Lines

The production and maintenance of TC-1 cells has been described previously. In brief, HPV-16 E6, E7 and *ras* oncogene were used to transform primary C57BL/6 mice lung epithelial cells to generate TC-1 cells. DC-1 cells were generated from the dendritic cell line provided by Dr. Kenneth Rock, University of Massachusetts. With continued passage, subclones of DCs

(DC-1) were generated that are easy to transfect (Kim, TW *et al.*, 2004, *Gene Ther. 11*:1011-1018). For the generation of TC-1/S and DC-1/S cells, the retroviral vector encoding the S protein of SARS-CoV was first generated. The phoenix packaging cells were transfected with pMSCV-S or pMSCV using Lipofectamine 2000. Supernatant from the transfected Phoenix (ϕNX) cells was incubated with 50% confluent TC-1 or DC-1 cells in the presence of polybrene (8 μg/ml; Sigma). Following transduction, the retroviral supernatants were removed from the transduced cells, and DCs were propagated in culture medium containing 7.5 μg/ml of puromycin for selection. The transduced TC-1 or DC-1 cells were further selected by growing in culture medium containing 10 μg/ml of puromycin for 5 days. The expression of S antigen was confirmed by Western blot analysis. All cells were maintained in RPMI medium (Invitrogen, Carlsbad, CA) supplemented with 2mM glutamine, 1mM sodium pyruvate, 20mM HEPES, 50μM β-mercaptoethanol, 100 IU/ml penicillin, 100μg/ml streptomycin and 10% fetal bovine serum (Gemini Bio-Products, Woodland, CA).

Western Blot Analysis

The expression of the full length protein S and its recombinant polypeptide fragments was examined in 293 cells transfected with various of the present DNA vectors encoding either no insert (control), S, S1, Si, S2, CRT or CRT/S1 was characterized by Western blot analysis. DNA, 20 μg, was transfected into 5×10^6 293 cells using lipofectamine® 2000 (Life Technologies, Rockville, MD). After overnight transfection, the cells were lysed with protein extraction reagent (Pierce, Rockford, IL). Equal amounts of proteins (50μg) were loaded and separated on a 10% SDS-PAGE gel. The gels were electroblotted onto a polyvinylidene difluoride membrane (Bio-Rad, Hercules, CA). Blots were blocked with PBS/0.05% Tween 20 (TTBS) containing 5% nonfat milk overnight at 4°C. Membranes were probed with rabbit anti-spike polyclonal antibody at 1:2000 dilution in TTBS for 1 hr at room temperature, washed six times with TTBS, and then incubated with goat anti-rabbit IgG conjugated to horseradish peroxidase (Zymed, San Francisco, CA) at 1:1000 dilution in TTBS containing 5% nonfat milk for 1 hr at room temperature. Membranes were washed four times with TTBS and developed using enhanced Hyperfilm-enhanced chemiluminescence (Amersham, Piscataway, NJ).

The presence of secreted S1 and CRT/S1 was confirmed by Western blot analysis. Forty eight hours after transfection as above with 20 μg of DNA encoding either no insert, S, S1, Si, S2, CRT or CRT/S1, 4 ml of culture supernatants were collected, centrifuged to remove cellular

debris and then was concentrated to 0.2 ml using an Amicon Ultra centrifugal filter device. Varying volumes (5, 10, 20 μl) of the concentrated supernatants were loaded and separated by SDS-10% PAGE before blotting. The presence of S polypeptides was detected by probing with Rabbit anti-S antibody at a 1:2000 dilution.

The presence of the S-specific antibody in sera from the mice immunized with the various DNA vaccines was determined by Western blot analysis using TC-1/S lysates as a source of antigen. The lysates from TC-1/No insert or TC-1/S were loaded and separated by SDS-10% PAGE gel before blotting. Immune serum samples were collected from DNA-vaccinated mice two weeks after the last vaccination and were diluted to 1:250 with PBS. Equal amounts of proteins (50 μg) from TC-1/No insert or TC-1/S lysates were probed with the diluted antisera from vaccinated mice.

Mice were as described in Example 1.

DNA Vaccination

DNA-coated gold particles were prepared and used as described above. C57BL/6 mice were immunized with 2 μg of the plasmid which included either no insert, S, S1, Si, S2, CRT or

Intracellular Cytokine Staining and Flow Cytometry Analysis

Using CD3 negative selection kit (Miltenyi Biotec, Auburn, CA), $CD3^+$ cells were enriched from splenocytes, harvested from mice one week after the last vaccination. DC cells (10^5) expressing S antigen (DC/S) were incubated with 10^6 of the isolated $CD3^+$ T cells for 16 hours. The DC cells not expressing S antigen (DC/No insert) served as a negative control. After activation, T cells were stained for both surface CD8 and intracellular IFN-γ, and analyzed with flow cytometry analysis as described before.

ELISA

The end-point dilution titer of S-specific antibodies in the sera from DNA-vaccinated C57BL/6 mice were determined by ELISA using 96 microwell plates coated with TC-1/S or TC-1/No insert cells. After overnight incubation, the cells (5×10^4/well) were washed once in phosphate buffered saline (PBS), then fixed and permeabilized using Cytofix/Cytoperm Kit (Pharmingen). Plates coated with cells were incubated with 1xPBS (0.3 ml/well) with 0.05% Tween 20 (PBT) containing 2% bovine serum albumin for 60 minutes at 37°C and washed again with PBT. Serial dilutions of the tested sera were added (0.1 ml/well) and the plates were

incubated for 60 minutes at 37°C. The plates were washed with PBT and were incubated with (0.1 ml/well) peroxidase-conjugated rabbit anti-mouse IgG (Zymed, San Francisco, CA) for 30 minutes at 37°C. The plates were washed with PT and incubated with (0.1 ml/well) peroxidase substrate according to the manufacturer's instructions for 15 minutes at 37 °C. Plates were read on a MicroElisa reader at a wavelength of 450 nm. Absorbance >3-fold above the absorbance from the negative controls were scored as positive reactions.

In Vivo Challenge with TC-1 cells expressing S antigen

The production and maintenance of TC-1 cells has been described previously. In brief, HPV-16 E6, E7 and *ras* oncogene were used to transform primary C57BL/6 mice lung epithelial cells to generate the TC-1 line.

For the construction of TC-1/S, supernatant from phoenix cells transfected with pMSCV-S was incubated with 50% confluent TC-1 cells in the presence of polybrene. The transduced TC-1 cells were further selected by growing in culture medium containing 10 μg/ml of puromycin for 5 days. The expression of S antigen was confirmed by Western analysis. For the challenge experiment, the immunized mice (10 per group) were subcutaneously challenged with 5×10^5 cells/mouse in the right leg one week after last vaccination, and then monitored twice a week to check the formation of TC-1/S tumor.

In vivo antibody depletion was performed to determine the contribution of various lymphocyte subsets to the protection, as described previously. The following mAbs were used: GK1.5 for CD4 depletion, mAb 2.43 for CD8 depletion, and mAb PK136 was used for NK1.1 depletion. Depletions were started one week after final vaccination. The immunized mice (10 per group) were challenged s.c. (5×10^5 cells/mouse) with TC-1/S cells one week after initiation of Ab depletion. The depletion was terminated on day 32 after challenge. The completeness of depletion was examined by flow cytometry. For each time point of analysis, >99% depletion of the appropriate subset was achieved while retaining normal levels of cells of the other subsets.

S-specific antibody responses

The presence of the S-specific antibody in sera from the mice immunized with the DNA vaccines encoding no insert, S, S1, Si, S2, CRT or CRT/S1 via a gene gun was detected by Western blot analysis. Immune serum samples were collected from DNA-vaccinated mice two weeks after the last vaccination and were diluted to 1:250 with PBS. Equal amounts of proteins (50 μg) from TC-1/No insert or TC-1/S lysates were probed with the diluted antisera. The end-

point dilution titer of S-specific antibodies in the sera from DNA-vaccinated C57BL/6 mice were determined by ELISA using 96 microwell plates coated with TC-1/S or TC-1/No insert cells. After overnight, the cells (5×10^4/well) were washed once in phosphate buffered saline (PBS), then fixed and permeabilized using Cyofix/Cytoperm Kit (Pharmingen). Plates coated with cells were incubated with (0.3 ml/well) PBS – 0.05% Tween 20 (PBT) containing 2% bovine serum albumin for 60 minutes at 37 °C and washed again with PBT. Serial dilutions of the tested sera were added (0.1 ml/well) and the plates were incubated for 60 minutes at 37 °C. The plates were washed with PBT and were incubated with (0.1 ml/well) peroxidase-conjugated rabbit anti-mouse IgG (Zymed, San Francisco, CA) for 30 minutes at 37 °C. The plates were washed with PT and incubated with (0.1 ml/well) peroxidase substrate (according to Sigma instructions) for 15 minutes at 37 °C. Plates were read on a MicroElisa reader at a wavelength of 450 nm. Reading higher than 3-fold negative controls were scored as positive reactions.

Statistical Analysis

All results expressed as means ± SD are representative of at least two different experiments. Data for intracellular cytokine staining with flow cytometry analysis and *in vivo* viral challenge experiments were evaluated by analysis of variance (ANOVA). Comparisons between individual data points were made using a student's *t*-test. In the tumor protection experiment, the principal outcome of interest was time to tumor development. The event time distributions for different mice were compared using the method of Kaplan and Meier and the log-rank statistic. $p< 0.05$ was considered significant.

Results

Cells transfected with the various S DNA immunogenic constructs expressed comparable levels of S protein

In order to characterize protein expression in cells (293 line) transfected with DNA constructs encoding the various domains of SARS-CoV S protein, Western blot analysis was done using rabbit anti-S polyclonal antibody. As shown in **Figure 7A**, lysates from 293 cells transfected with the various DNA constructs revealed protein bands correlated with the expected sizes of S, S1, Si and S2. Furthermore, levels of protein expression by 293 cells transfected with the various DNA constructs appeared to be comparable. As shown in **Figure 7B**, only cells transfected with the S1 DNA construct were able to secrete S1 protein. In contrast, cells transfected with S, Si or S2 DNA did not secrete the encoded proteins.

DNA encoding S1 generates the highest S-specific antibody immune response in vaccinated mice.

To determine the antibody immune response induced by immunization with the various DNA constructs encoding the domains of S protein, a study was done in which mice received pcDNA3-S, pcDNA3-S1, PcDNA3, Si, pcDNA3-S2 or pcDNA3. Two weeks after the last booster, sera were collected and antibodies against S protein were measured. TC-1/S cell lysates were used as a source of S protein for Western Blot analysis as well as for ELISA. Figure 8A shows that sera diluted 1:250 as probes in Western blots revealed that mice given the S1 DNA construct generated the highest S-specific antibody immune response. Immunization with DNA encoding the full length S protein also resulted in an S-specific antibody responses, albeit lower. Similar results were observed when testing these sera in ELISA. As shown in Figure 8B, mice given S1 DNA generated the greatest S-specific antibody responses. Thus, administration of DNA that encodes the receptor-binding domain (S1) of SARS-CoV S protein is capable of generating stronger S-specific antibody responses than does administration of DNA encoding the full length S protein. S1 is therefore an excellent target for development of preventive SARS-CoV DNA vaccines of the type disclosed herein.

Vaccination with DNA encoding SARS CoV S1 generates the higher numbers of S-specific $CD8^+$ T cells *in vivo*

To assess the numbers of S-specific $CD8^+$ T-cell precursors that are triggered following administration of various of the DNA constructs to mice, intracellular cytokine staining was done in conjunction with flow cytometric analysis using $CD3^+$ cells enriched T cells from spleens of vaccinated mice one week after the last vaccination. Enriched $CD3^+$ T cells enriched cells from immunized mice were stimulated *in vitro* with DCs transfected with DNA encoding SARS CoV S protein (or as a control, DNA without an insert). After overnight incubation, cells were stained for both CD8 and intracellular IFNγ. As shown in **Figure 9A and 9B**, pcDNA3-S1 induced the highest number of S-specific $IFN\gamma^+$ $CD8^+$ T-cell precursors among all the DNA constructs tested ($p<0.01$). Vaccination with pcDNA3-S or pcDNA3-Si also induced S-specific $CD8^+$ T cells to a larger extent that did pcDNA3-S2 ($p<0.05$), but less than did S1 DNA. These results indicate that pcDNA3-S1 is the more potent immunogen for S-specific $CD8^+$ T cell immune responses. Taken together, the results argue in favor or the receptor binding domain of SARS CoV S protein represents as a desirable target for generating SARS-CoV S specific antibodies as well as CD8+ T cell reactivity (likely cytotoxic T cells)..

Cells transfected with the DNA encoding calreticulin linked to S1 generate comparable levels of S protein as DNA encoding S1.

Some of th present inventors identified the use of DNA constructs comprising sequences encoding calreticulin (CRT) as an excellent strategy to enhance antigen-specific and T cell mediated immune responses to DNA vaccines that comprise DNA encoding an antigen. In the present, a DNA construct was made that encoded CRT linked to S1.

Expression of such DNA was tested by transfecting 293 cells with the DNA constructs and performing Western blot analysis using rabbit anti-S polyclonal antibody. As shown in **Figure 10A**, lysates from 293 cells transfected with the CRT/S1 or S1 DNA revealed protein bands correlated with the expected sizes of the fusion polypeptide CRT/S1 or of S1 alone. Furthermore, the level of protein expression by 293 cells transfected with the these DNA constructs appeared to be comparable. As shown in **Figure 10B**, cells transfected with CRT/S1 DNA and with S1 DNA construct could secrete S1 protein.

DNA encoding CRT/S1 is a potent stimulator of S-specific antibody responses in vaccinated mice

Mice were immunized with pcDNA3-CRT/S1, pcDNA3-S1, PcDNA3-CRT or pcDNA3. Two weeks after the last booster, sera were collected and assayed for antibodies against S protein. TC-1/S cell lysates were used as a source of S protein for Western Blot analysis as well as in ELISA. As shown in **Figure 11A**, examining sera diluted at 1:250 in Western blot analysis, it was found that mice vaccinated with the CRT/S1 DNA generated the highest S-specific antibody response. Vaccination with DNA encoding S1 also generated S-specific antibody responses, albeit lower than vaccination with the CRT/S1 construct. ELISA gave similar results in characterizing the S-specific antibody response. As shown in Figure 11B, mice vaccinated with CRT/S1 DNA generated the highest S-specific antibody response. Thus, vaccination with DNA encoding CRT linked to a SARS antigen, the receptor-binding domain (S1) of SARS-CoV S protein, generated enhanced S-specific antibody responses vs vaccination with DNA encoding the S1 protein alone.

Vaccination with DNA encoding CRT/S1 stimulates S-specific $CD8^+$ T cells in vaccinated mice

To assess the quantity of S-specific $CD8^+$ T-cell precursors generated by administration of the various DNA S protein constructs (pcDNA3-CRT/S1, pcDNA3-S1, PcDNA3-CRT or empty pcDNA3), intracellular cytokine staining was performed with flow cytometric analysis using $CD3^+$ T cells enriched from spleens of vaccinated mice one week after the last

vaccination. These T cells were stimulated *in vitro* with DCs transfected with DNA encoding S protein or control DNA, and stained for both CD8 and intracellular IFNγ. As shown in **Figure 12A and 12B**, vaccination with pcDNA3-CRT/S1 was the most potent in generating S-specific $IFN\gamma^+$ $CD8^+$ T-cell (compared to vaccination with pcDNA3-S1) ($p < 0.005$). Vaccination with either of the two controls (pcDNA3-CRT or pcDNA3) resulted in only background levels of S-specific $CD8^+$ T cells. These results indicate that vaccination with pcDNA3-CRT/S1 chimeric construct generates higher numbers of antigen-specific $CD8^+$ T cells *in vivo* compared to vaccination with pcDNA3-S1. Thus, in addition to some of the present inventors' successes using the CRT strategy with human papillomavirus vaccines (the E6 and E7 protein; see, for example WO02/012281) the present results show that S1 DNA vaccines employing the CRT strategy are potent in generating SARS-CoV S specific humoral and CD8+ T cell-mediated immune responses.

Vaccination with DNA encoding CRT/S1 is generates preventive antitumor immunity against tumor cells that are engineered to express the SARS CoV S protein

A non-infectious model system was employed to determine a therapeutic outcome of the immunity generated by the present constructs and the enhancing effect of the CRT DNA on such immunity. An antitumor response was examined using an *in vivo* tumor protection assay. TC-1/S tumor cells, transfected to express the S protein were the target of the immunity. As shown in **Figur 13A**, 100% of mice receiving CRT/S1 DNA remained tumor-free 35 days after TC-1/S challenge. Ir comparison, only 40% of the mice receiving S1 DNA remained tumor-free at this time. All mice vaccinated with control CRT constructs or pcDNA3 plasmid controls grew tumors within two weeks after challenge.

To confirm which subsets of lymphocytes were important for this therapeutic effect, an *in viv* antibody depletion study was conducted. Its results appear in **Figure 8B.** All mice depleted of CD8 cells grew tumors within 10 days after TC-1/S challenge. In contrast, 100% of mice depleted of CD cells or NK cells remained tumor-free 35 days after challenge. Thus, $CD8^+$ T cells are required for t therapeutic (antitumor) effect of the CRT/S1 DNA vaccine. Thus, the T cell-mediated immunity generated by immunization or vaccination with CRT/S1 DNA can effect clinical-type therapeutic results, measured here as an antitumor effect.

EXAMPLE 3

<u>DNA Vaccines Targeting the Membrane Protein (M) of SARS-CoV</u>

Materials and Methods

Plasmid DNA Constructs and DNA Preparation

In the current study we used the mammalian expression vector, pcDNA3.1/myc-His (-) (Invitrogen, Carlsbad, CA) for our DNA vaccine studies. For the generation of pcDNA3-M-myc, the DNA fragment encoding SARS-Co V membrane antigen (M) was amplified with PCR using a set of primers:

5'-aaagaattcatggcagacaacggtactattac-3', SEQ ID NO:50
5'-tttggtaccttactgtactagcaaagcaatat-3' SEQ ID NO:51

and pGEX-1-MG6 as a template. The amplified product was further cloned into the EcoRI/KpnI sites of pcDNA3.1/*myc*-His (-) vector. To generate pcDNA3-CRT-myc, CRT DNA segment was isolated from pcDNA3-CRT and cloned into the XhoI/EcoRI sites of pcDNA3.1/*myc*-His(-). For the generation of pcDNA3-CRT/N-myc, the amplified M DNA was cloned into the EcoRI/KpnI sites of pcDNA3-CRT-myc. The accuracy of these constructs was confirmed by DNA sequencing. The DNA was amplified in E. coli DH5α and purified as described previously .

Cell Lines: Construction of DC expressing M

The production and maintenance of TC-1 cells and DC-1 cells was described above. To generate SARS CoV membrane antigen presenting cell, the immortalized DC line, which was kindly provided by Dr. Kenneth Rock (University of Massachusetts, Worcester, MA), was genetically manipulated by retroviral system. For this, the cDNA of M was isolated from pGEX-1-MG6 after BamHI/EcoRI restriction and further cloned into the BglII/EcoRI sites of pMSCV vector (Invitrogen). Phoenix (ϕNX) packaging cells were transfected with pMSCV-M or pMSCV using Lipofectamine 2000. Supernatants from the transfected phoenix cells were incubated with 50% confluent DC in the presence of polybrene (8ug/ml; Sigma). Following transduction, the retroviral supernatants were removed, and DCs were propagated in culture medium containing 7.5 μg/ml of puromycin for selection. The expression of M antigen was confirmed by western blot analysis.

For the generation of TC-1/M and DC-1/M cells, we first generate retroviral vector encoding the M protein of SARS-CoV. The phoenix packaging cells were transfected with pMSCV-M or pMSCV using Lipofectamine 2000. Supernatant from the transfected Phoenix (ϕNX) cells was incubated with 50% confluent TC-1 or DC-1 cells in the presence of polybrene

(8 μg/ml; Sigma). Following transduction, the retroviral supernatants were removed from the transduced cells, and DCs were propagated in culture medium containing 7.5 μg/ml of puromycin for selection. The transduced TC-1 or DC-1 cells were further selected by growing in culture medium containing 10 μg/ml of puromycin for 5 days. The expression of M antigen was confirmed by Western blot analysis. All cells were maintained in supplemented RPMI medium as above.

Western Blot Analysis

The expression of M protein in TC-1/M, DC-1/M or 293 cells transfected with pcDNA3.1/*myc*-His (-) encoding no insert, CRT, M, or CRT/M DNA was characterized by Western blot analysis. 5×10^6 293 cells were transfected with 20 μg of DNA using lipofectamine 2000 (Life Technologies, Rockville, MD). The remaining methods were as in the previous Examples.

Mice – were as described above.

DNA Vaccination

DNA-coated gold particles were prepared and used as described above. C57BL/6 mice were immunized with 2 μg of the plasmid encoding no insert, CRT, M, or CRT/M protein.

Intracellular Cytokine Staining and Flow Cytometry Analysis

This was described above. DC cells expressing M antigen (DC/M), 10^5 were incubated with 10^6 isolated $CD3^+$ T cell for 16 hours. The DC cells not expressing M antigen (DC/No insert) were used as a negative control. After activation, T cells were stained for surface CD8 or CD4 and intracellular IFNγ or IL-4 and analyzed flow cytometrically as described.

In Vivo Challenge with TC-1 expressing M antigen

The production and maintenance of TC-1 cells has been described previously.

For the construction of TC-1/M cells, supernatant from the transfected phoenix cells with pMSCV-M was incubated with 50% confluent TC-1 as described in the earlier Examples. The expression of M antigen was confirmed by Western blot. Tumor Challenge experiments were as above. *In vivo* antibody depletions was performed as above.

Statistical Analysis – as above

RESULTS

Cells transfected with M or CRT/M DNA vaccines generate comparable levels of M protein.

In order to characterize M protein expression in cells (293 line) transfected with DNA constructs encoding SARS-CoV M or CRT/M, Western blot analysis was done using mouse anti-Myc antibody. 293 cells transfected with DNA encoding CRT or DNA without insert were used as controls. As shown in **Figure 14**, lysates from cells transfected with the various DNA constructs revealed protein bands having the expected sizes of M and CRT/M. 293 cells transfected with M and CRT/M DNA vaccines expressed comparable levels of the encoded proteins.

Vaccination with DNA encoding CRT/M generates higher numbers of M-specific CD8$^+$ T cells *in vivo*

To assess the quantity of M-specific CD8$^+$ T-cell precursors generated by the pcDNA3, pcDNA3-CRT, pcDNA3-M or pcDNA3-CRT/M vaccine constructs in vaccinated mice,

To assess the numbers of M-specific CD8$^+$ T-cell precursors that are triggered following administration of various of the DNA constructs (pcDNA3 control, pcDNA3-CRT control, pcDNA3-M and pcDNA3-CRT/M) to mice, intracellular cytokine staining was done in conjunction with flow cytometric analysis using spleen cells from the vaccinated mice one week after the last vaccination. Pooled spleen cells were stimulated *in vitro* with DCs transfected with DNA encoding M protein or, as a control, DNA with no insert and stained for both CD8 and intracellular IFNγ. As shown in **Figure 15A and 15B**, pcDNA3-CRT/M induced the highest number of M-specific IFNγ$^+$ CD8$^+$ T-cell precursors when compared to pcDNA3-M ($p < 0.005$). Vaccination with pcDNA3-CRT or pcDNA3 only generated background levels of M-specific CD8$^+$ T cells. These results indicate that vaccination with pcDNA3-CRT/M is the more potent immunogen for M-specific CD8$^+$ T cells immune responses. Thus M protein DNA vaccines employing the CRT strategy are effective in stimulating strong SARS-CoV M-specific CD8+ T cell reactivity (likely to include cytotoxic T cells).

Vaccination with DNA encoding CRT/M generates high numbers of M-specific CD4$^+$ T helper cells

To assess the numbers of M-specific CD4$^+$ T cells generated by the same DNA constructs, intracellular cytokine staining and flow cytometric analysis was done on spleen cells from vaccinated mice harvested one week after the last vaccination. Pooled cells were stimulated *in vitro* with DCs transfected with DNA encoding M protein or, as a control, DNA with no insert. After overnight incubation, cells were stained for both CD4 and intracellular IFNγ or IL-4. As shown in **Figure 16A and 16B,** pcDNA3-CRT/M induced the higher number of M-

specific IFNγ^+ CD4$^+$ T helper type 1 (Th1) cells compared to pcDNA3-M ($p<0.005$). Control vaccination (pcDNA3-CRT or pcDNA3) generated only background levels of M-specific CD4$^+$ Th1 cells. These results further support the success of the CRT strategy in generating greater numbers of M-specific CD4$^+$ Th1 as compared to immunization with DNA encoding antigen alone (*e.g.*, pcDNA3-M).

IL-4-secreting M-specific CD4$^+$ T helper cells of the Th2 class were measured after administering the two experimental and two control DNA vaccine preparations as assessed by intracellular cytokine staining followed by flow cytometric analysis. As shown in **Figure 17A and 17B**, vaccination with pcDNA3-CRT/M triggered higher numbers of IL-4-secreting M-specific CD4$^+$ T cells compared to pcDNA3-M (p value<0.05), although the absolute numbers of IL-4-secreting M-specific CD4$^+$ T cells was lower than the number of IFNγ-secreting, M-specific CD4$^+$ Th1 cells in CRT/M-vaccinated mice. The two control plasmids, pcDNA3-CRT and pcDNA3 resulted in only background levels of M-specific CD4$^+$ Th2 cells. Taken together, the results indicate that M DNA vaccines employing the CRT strategy are potent stimuli for SARS-CoV M-specific IFN-γ-secreting, CD4$^+$ and CD8$^+$ T cells.

Immunization with pcDNA3-CRT/M generates protective antitumor immunity against tumor cells that are engineered to express the SARS CoV M protein.

As discussed in Example 2, a non-infectious model system was employed to determine a therapeutic outcome of the immunity generated by the present constructs and the enhancing effect of the CRT DNA on such immunity. An antitumor response was examined using an *in vivo* tumor protection assay. TC-1/M tumor cells, transfected to express the M protein, were the target of the immunity. As shown in **Figure 18A**, 100% of mice receiving pcDNA3-CRT/M remained tumor-free six weeks after TC-1/M challenge. In contrast, all animals vaccinated with the control plasmid (no insert) or the pcDNA3-CRT plasmid, developed tumors within 10 days after the tumor challenge. Therefore, the CRT/M DNA construct was capable of generating not only a high number of M-specific T cells *in vitro* but also a protective antitumor effect against challenge with M-expressing tumor cells in vaccinated mice.

To confirm which subsets of lymphocytes were important for this therapeutic effect, an *in vivo* antibody depletion study was conducted. Its results appear in **Figure 18B.** All mice depleted of CD8$^+$ T cells grew tumors within 15 days of TC-1/M challenge. In contrast, 100% of mice depleted of CD4$^+$ T cells or NK cells remained tumor-free. Thus, CD8$^+$ T cells are required for the therapeutic (antitumor) effect of the CRT/S1 DNA vaccine. Thus, the T cell-mediated immunity generated by

immunization or vaccination with CRT/S1 DNA can effect clinical-type therapeutic results, measured here as an antitumor effect.

The references cited above are all incorporated by reference herein, whether specifically incorporated or not.

Having now fully described this invention, it will be appreciated by those skilled in the art that the same can be performed within a wide range of equivalent parameters, concentrations, and conditions without departing from the spirit and scope of the invention and without undue experimentation. While this invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth as follows in the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A nucleic acid molecule encoding a fusion polypeptide useful as a vaccine composition, which molecule comprises:

(a) a first nucleic acid sequence encoding a first polypeptide that comprises an endoplasmic reticulum chaperone polypeptide;

(b) optionally, fused in frame with the first nucleic acid sequence, a linker nucleic acid sequence encoding a linker peptide; and

(c) a second nucleic acid sequence that is linked in frame to said first nucleic acid sequence or to said linker nucleic acid sequence and that encodes an antigenic polypeptide or peptide from a SARS-CoV,

said SARS-CoV antigenic polypeptide or peptide being one that is the target of a protective or neutralizing immune response.

2. The nucleic acid molecule of claim 1, wherein the antigenic peptide comprises an epitope that binds to a MHC class I protein.

3. The nucleic acid molecule of claim 2, wherein said epitope is between about 8 amino acid residues and about 11 amino acid residues in length.

4. The nucleic acid molecule of claim 1 wherein the chaperone polypeptide comprises calreticulin or an immunologically active fragment or variant thereof.

5. The nucleic acid molecule of claim 4, wherein said calreticulin is human calreticulin having the amino acid sequence SEQ ID NO:2 and wherein the active fragment or variant is a fragment or variant of SEQ ID NO:2.

6. The nucleic acid molecule of claim 4, wherein the first nucleic acid sequence comprises the coding portion of SEQ ID NO:1, or of a fragment or variant thereof.

7. The nucleic acid molecule of claim 5 wherein the calreticulin consists essentially of a sequence from about residue 1 to about residue 180 of SEQ ID NO:2.

8. The nucleic acid molecule of claim 5, wherein the calreticulin consists essentially of a sequence from about residue 181 to about residue 417 of SEQ ID NO:2.

9. The nucleic acid molecule of claim 1, wherein the chaperone polypeptide comprises

(a) a calnexin polypeptide or an equivalent thereof;

(b) an ER60 polypeptide or an equivalent thereof;

(c) a tapasin polypeptide or an equivalent thereof; or

(d) a GRP94/GP96 polypeptide, a GRP9 4 polypeptide or an equivalent thereof.

10. The nucleic acid molecule of any of claims 1-9 wherein the antigen is one which is present on, or cross-reactive with an epitope of a SARS-CoV structural protein

11. The nucleic acid molecule of claim 10 wherein the antigen is from a strain or isolate of SARS-CoV selected from the group consisting of TOR2 and TW1.

12. The nucleic acid molecule of claim 10 wherein the structural protein is selected from the group consisting of the Spike (S) protein, the envelope (E) protein, the membrane (M) protein, and the nucleocapsid (N) protein.

13. The nucleic acid molecule of claim 10 wherein the structural protein is the S protein having an amino acid sequence SEQ ID NO:14 or a domain or fragment thereof.

14. The nucleic acid molecule of claim 13 wherein the domain or fragment is selected from the group consisting of SEQ ID NO:15, SEQ ID NO:16 and SEQ ID NO:17

15. The nucleic acid molecule of claim 10 wherein the structural protein is the E protein having an amino acid sequence SEQ ID NO:19 or a fragment thereof.

16. The nucleic acid molecule of claim 10 wherein the structural protein is the M protein having an amino acid sequence SEQ ID NO:21 or a fragment thereof.

17. The nucleic acid molecule of claim 10 wherein the structural protein is the N protein having an amino acid sequence SEQ ID NO:23 or a fragment thereof.

18. The nucleic acid molecule of claim 10 having a sequence selected from the group consisting of SEQ ID NO:24, SEQ ID NO:27 or SEQ ID NO:30.

19. An expression vector or cassette comprising the nucleic acid molecule of any of claims 1-9 operatively linked to

(a) a promoter; and

(b) optionally, additional regulatory sequences that regulate expression of said nucleic acid in a eukaryotic cell.

20. The expression vector or cassette of claim 19 wherein the antigen is one which is present on, or cross-reactive with an epitope of a SARS-CoV structural protein.

21. The expression vector or cassette of claim 20 wherein the structural protein is selected from the group consisting of the Spike (S) protein, the envelope (E) protein, the membrane (M) protein, and the nucleocapsid (N) protein.

22. The expression vector or cassette of claim 20 which is a viral vector or a plasmid.

23. The expression vector or cassette of claim 20, wherein the chaperone polypeptide comprises a calreticulin polypeptide or an active fragment thereof.

24 The expression vector or cassette of claim 23 wherein the calreticulin polypeptide:

(i) comprises amino acid sequence SEQ ID NO:2 ; or

(ii) is encoded by the coding portion of the nucleic acid molecule having the sequence SEQ ID NO:1.

25. The expression vector or cassette of claim 20, wherein the chaperone polypeptide comprises any one or more of a tapasin, an ER60, an ERP94 or a calnexin polypeptide, or an equivalent thereof.

26. A cell which has been modified to express the nucleic acid molecule of any of claims 1-9.

27. A cell which has been modified to comprise the expression vector or cassette of claim 19.

28. A particle suitable for introduction into a cell or an animal by particle bombardment comprising the nucleic acid of any of claims 1-9.

29. A particle suitable for introduction into a cell or an animal by particle bombardment comprising expression cassette or vector of any of claims 20.

30. The particle of claim 29 wherein the particle comprises gold.

31. A fusion or chimeric polypeptide comprising

(a) a first polypeptide comprising an endoplasmic reticulum chaperone polypeptide; and

(b) a second polypeptide comprising an antigenic polypeptide or peptide from a SARS-CoV,

said SARS-CoV antigenic polypeptide or peptide being one that is the target of an anti-viral immune response.

32. The fusion or chimeric polypeptide of claim 31 wherein the chaperone polypeptide comprises a calreticulin polypeptide, an active fragment thereof, or a homologue thereof.

33. The fusion or chimeric polypeptide of claim 32 wherein the calreticulin polypeptide is a human calreticulin polypeptide that::

(i) comprises amino acid sequence SEQ ID NO:2 ; or

(ii) is encoded by a coding portion of the nucleic acid molecule having the sequence SEQ ID NO:1.

34. The fusion or chimeric polypeptide of claim 31, wherein the antigenic peptide or polypeptide corresponds to a SARS-CoV structural protein is a selected from the group consisting of the Spike (S) protein, the envelope (E) protein, the membrane (M) protein, and the nucleocapsid (N) protein.

35. The fusion or chimeric polypeptide of claim 31wherein the chaperone polypeptide and the antigenic polypeptide or peptide are linked by a chemical linker.

36. The fusion polypeptide of any of claims 31-35 wherein the first polypeptide is N-terminal to the second polypeptide.

37. The fusion polypeptide of any of claims 31-35 wherein the second polypeptide is N-terminal to the first polypeptide.

38. The fusion or chimeric polypeptide of claim 31 wherein the chaperone polypeptide comprises any one or more of a tapasin, an ER60, an ERP94 or a calnexin polypeptide, or an equivalent thereof.

39. A pharmaceutical composition capable of inducing or enhancing a SARS-CoV antigen-specific immune response, comprising:

(a) pharmaceutically and immunologically acceptable excipient in combination with;

(b) the nucleic acid molecule of claim 1-9.

40. A pharmaceutical composition capable of inducing or enhancing a SARS-CoV antigen-specific immune response, comprising:

(a) pharmaceutically and immunologically acceptable excipient in combination with;

(b) the expression vector or cassette of claim 19.

41. A pharmaceutical composition capable of inducing or enhancing a SARS-CoV antigen-specific immune response, comprising:

(a) pharmaceutically and immunologically acceptable excipient in combination with;

(b) the expression vector or cassette of claim 20.

42. A pharmaceutical composition capable of inducing or enhancing a SARS-CoV antigen-specific immune response, comprising:

(a) pharmaceutically and immunologically acceptable excipient in combination with;

(b) the expression vector or cassette of claim 21.

43. A pharmaceutical composition capable of inducing or enhancing a SARS-CoV antigen-specific immune response, comprising:

(a) pharmaceutically and immunologically acceptable excipient in combination with;

(b) the fusion or chimeric polypeptide of claim 31.

44. A pharmaceutical composition capable of inducing or enhancing a SARS-CoV antigen-specific immune response, comprising:

(a) pharmaceutically and immunologically acceptable excipient in combination with;

(b) the particle of claim 29.

45. A method of inducing or enhancing a SARS-CoV antigen specific immune response in a subject comprising administering to the subject an effective amount of the pharmaceutical composition of claim 39, thereby inducing or enhancing said response.

46. A method of inducing or enhancing a SARS-CoV antigen specific immune response in a subject comprising administering to the subject an effective amount of the pharmaceutical composition of claim 40, thereby inducing or enhancing said response.

47. A method of inducing or enhancing a SARS-CoV antigen specific immune response in a subject comprising administering to the subject an effective amount of the pharmaceutical composition of claim 41, thereby inducing or enhancing said response.

48. A method of inducing or enhancing a SARS-CoV antigen specific immune response in a subject comprising administering to the subject an effective amount of the pharmaceutical composition of claim 42, thereby inducing or enhancing said response.

49. A method of inducing or enhancing a SARS-CoV antigen specific immune response in a subject comprising administering to the subject an effective amount of the pharmaceutical composition of claim 43, thereby inducing or enhancing said response.

50. A method of inducing or enhancing a SARS-CoV antigen specific immune response in a subject comprising administering to the subject an effective amount of the pharmaceutical composition of claim 44, thereby inducing or enhancing said response.

51. The method of claim 45, wherein the response is mediated at least in part by $CD8^+$ cytotoxic T lymphocytes (CTL).

52. The method of claim 45, wherein the response is mediated at least in part by antibodies.

53. The method of claim 45 wherein said administering is by a intramuscular, intradermal, or subcutaneous route.

54. The method of claim 45 wherein administering is by biolistic injection of said nucleic acid molecule.

55. A method of inducing or enhancing an antigen specific lymphocyte response or immune response in cells or in a subject comprising providing to said cells or to said subject an effective amount of the pharmaceutical composition of claim 39-44, thereby inducing or enhancing said response.

56. A method of increasing the numbers or lytic activity of $CD8^+$ T cells specific for a selected SARS-CoV antigen in a subject, comprising administering to said subject an effective amount of the pharmaceutical composition of claim 45, wherein

(i) said nucleic acid molecule encodes said selected antigen, and

(ii) said selected SARS-CoV antigen comprises an epitope that binds to, and is presented on the cell surface by, MHC class I proteins,

thereby increasing the numbers or activity of said CTLs.

57. A method of inhibiting a viral infection by a SARS-CoV or preventing or diminishing spread of said virus in a subject, comprising administering to said subject an effective amount of a pharmaceutical composition of claim 45, wherein said nucleic acid molecule encodes one or more SARS-CoV epitopes present on said virus or on virus infected cells in said subject, thereby inhibiting said infection or preventing or diminishing said spread.

58. The method of claim 57, further comprising before, together with or after said administration of said pharmaceutical composition, administering to said subject a second composition having effective SARS-CoV-directed anti-viral activity.

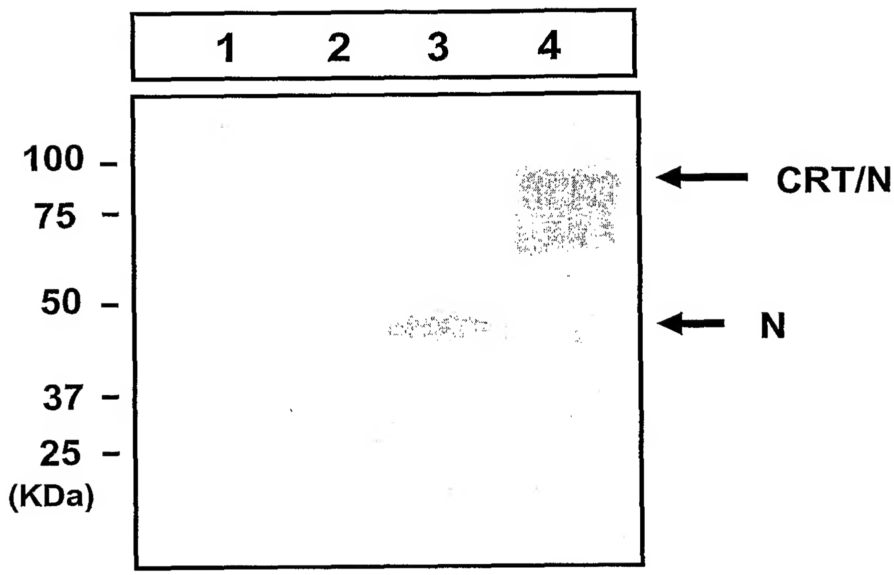


Fig. 1

Fig. 2A

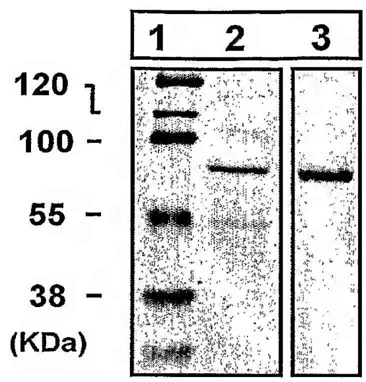
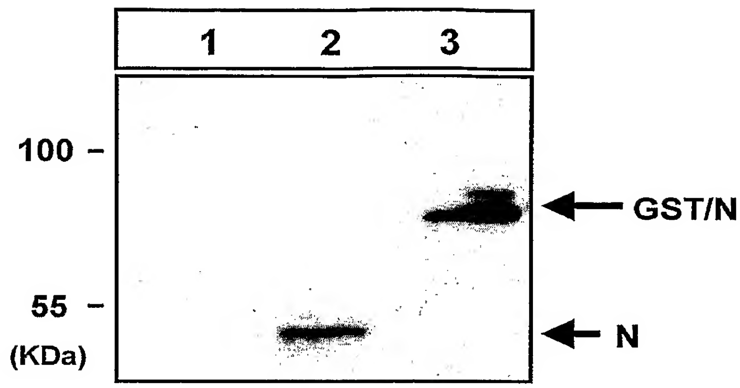


Fig. 2B



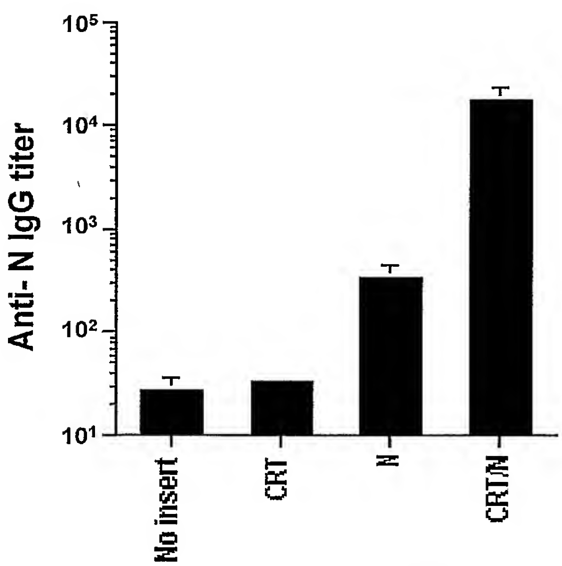


Fig. 2C

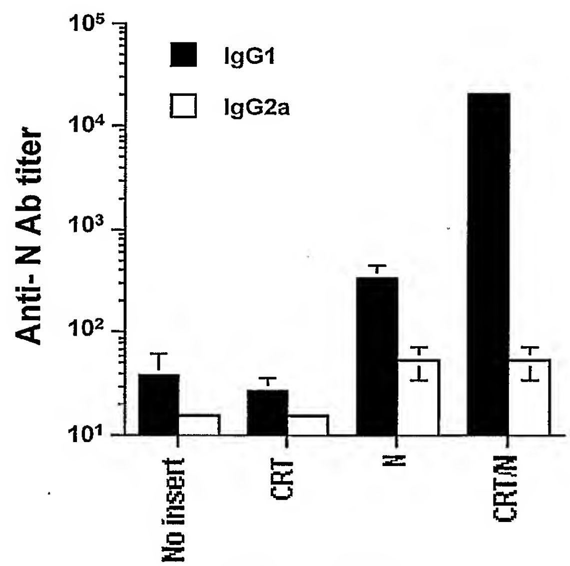


Fig. 2D

Fig. 3A

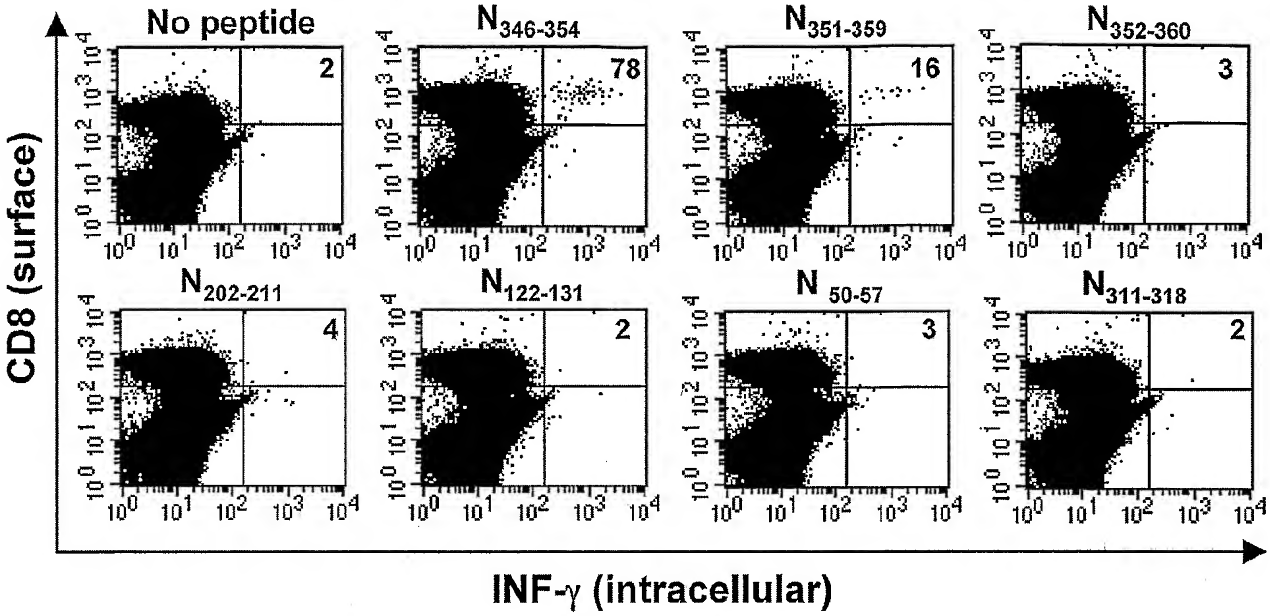


Fig. 3B

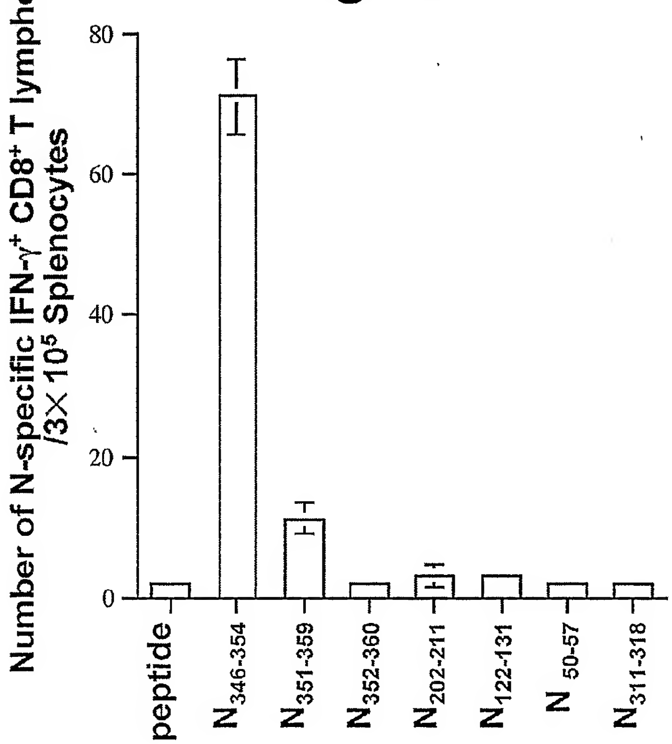


Fig. 3C

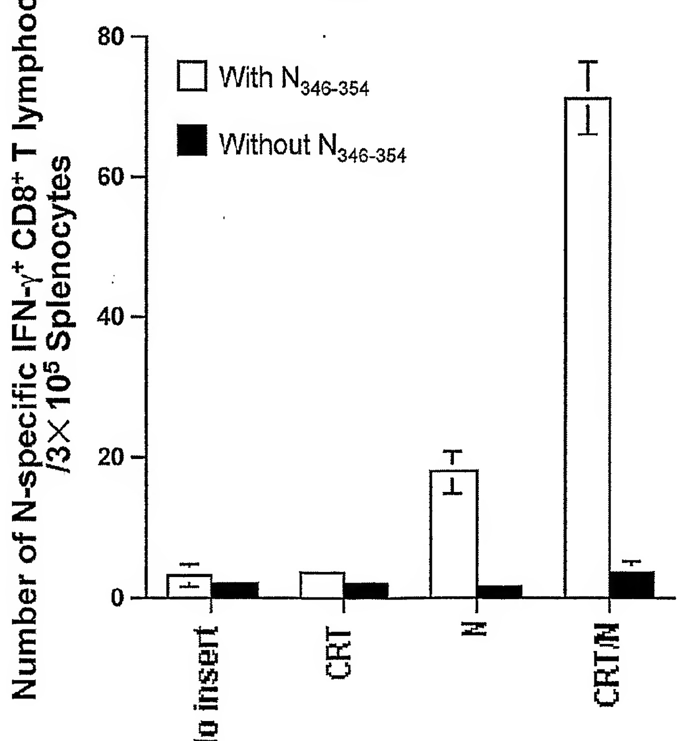
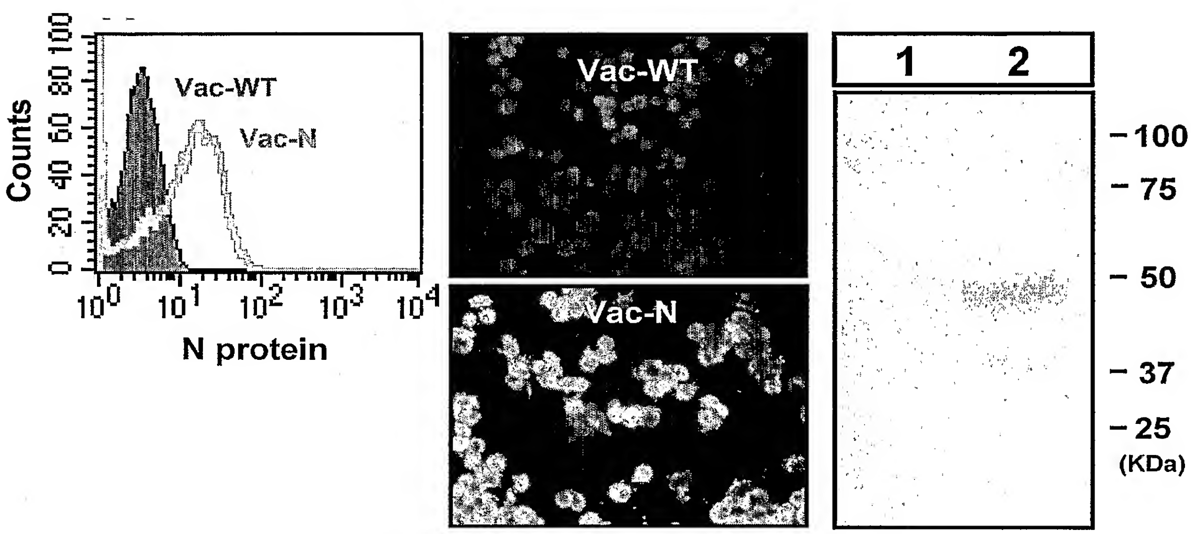


Fig 4A

Fig 4B

Fig 4C



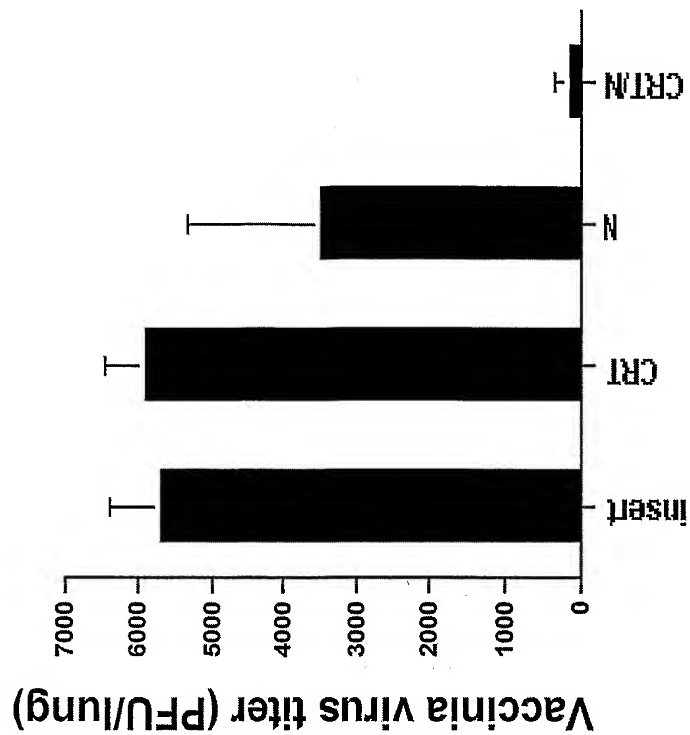

Vaccinia virus titer (PFU/lung)
7000
6000
5000
4000
3000
2000
1000
0
No insert
CRT
N
CRT/N

Fig. 5B

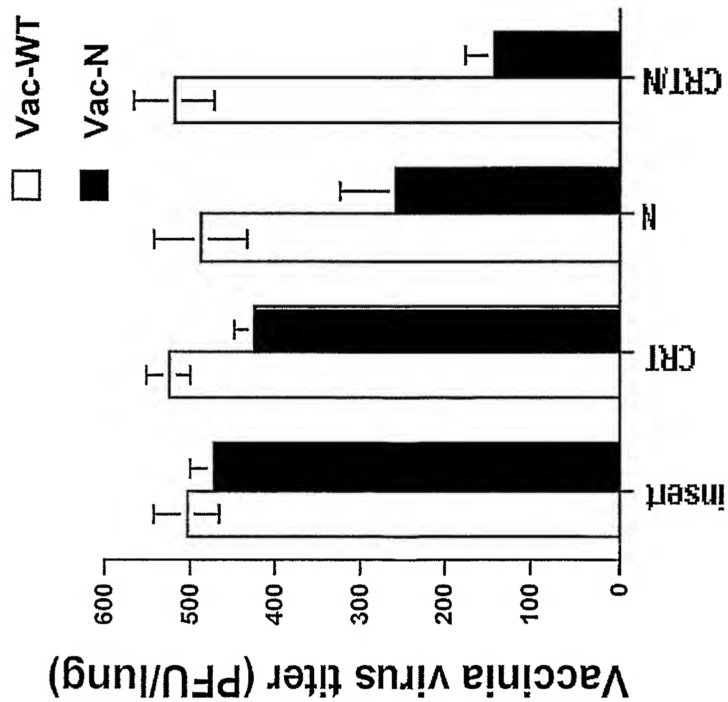

Vaccinia virus titer (PFU/lung)
600
500
400
300
200
100
0
No insert
CRT
N
CRT/N
Vac-WT
Vac-N

Fig. 5A

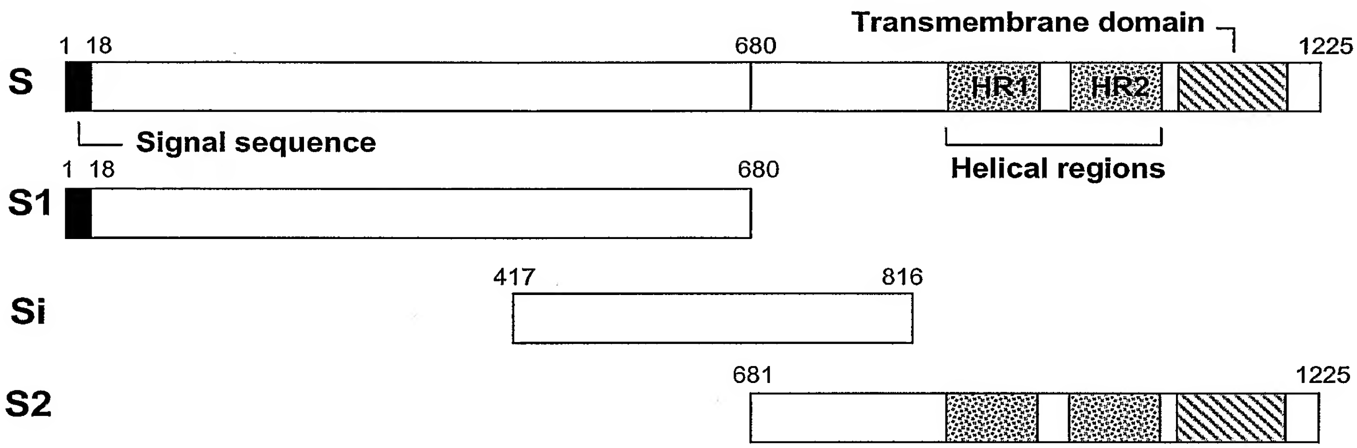

Transmembrane domain
1 18
680
1225
S
HR1
HR2
Signal sequence
Helical regions
1 18
680
S1
417
816
Si
681
1225
S2

Fig. 6

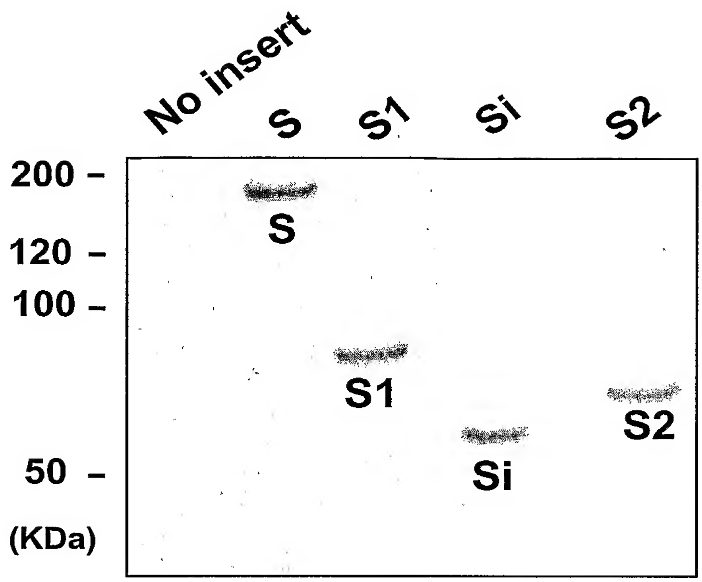

No insert
S
S1
Si
S2
200 –
120 –
100 –
50 –
(KDa)
S
S1
Si
S2

Fig. 7A

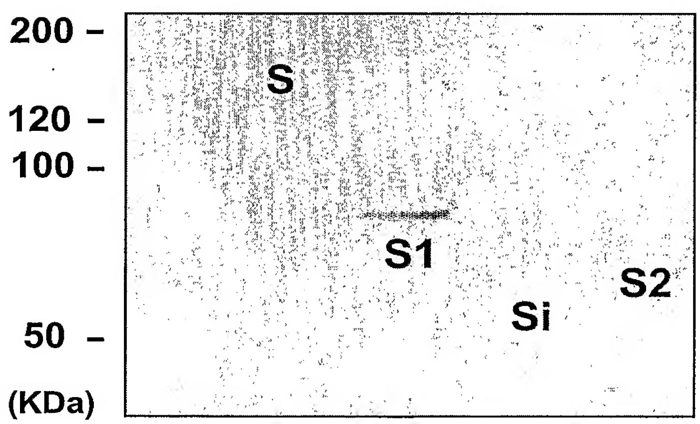

200 –
120 –
100 –
50 –
(KDa)
S
S1
Si
S2

Fig. 7B

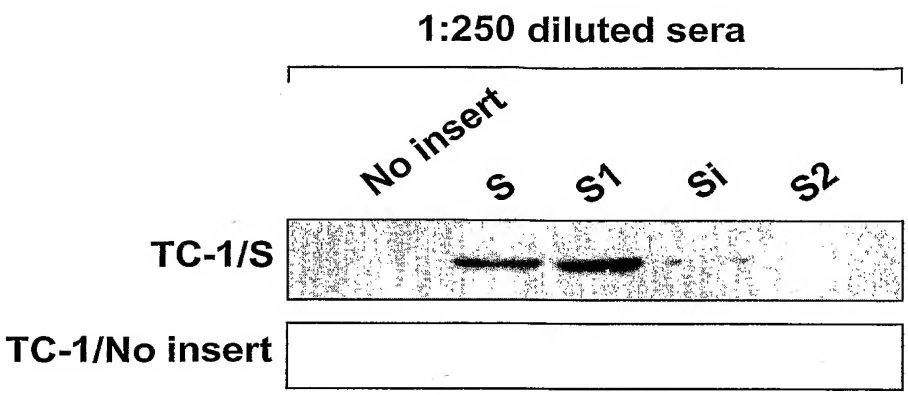

1:250 diluted sera
No insert
S
S1
Si
S2
TC-1/S
TC-1/No insert

Fig. 8A

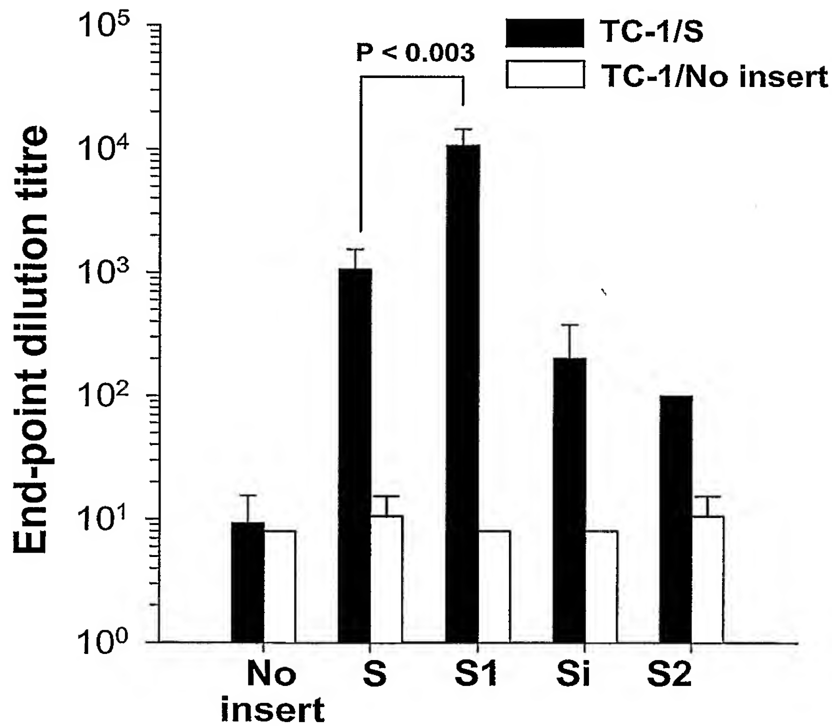
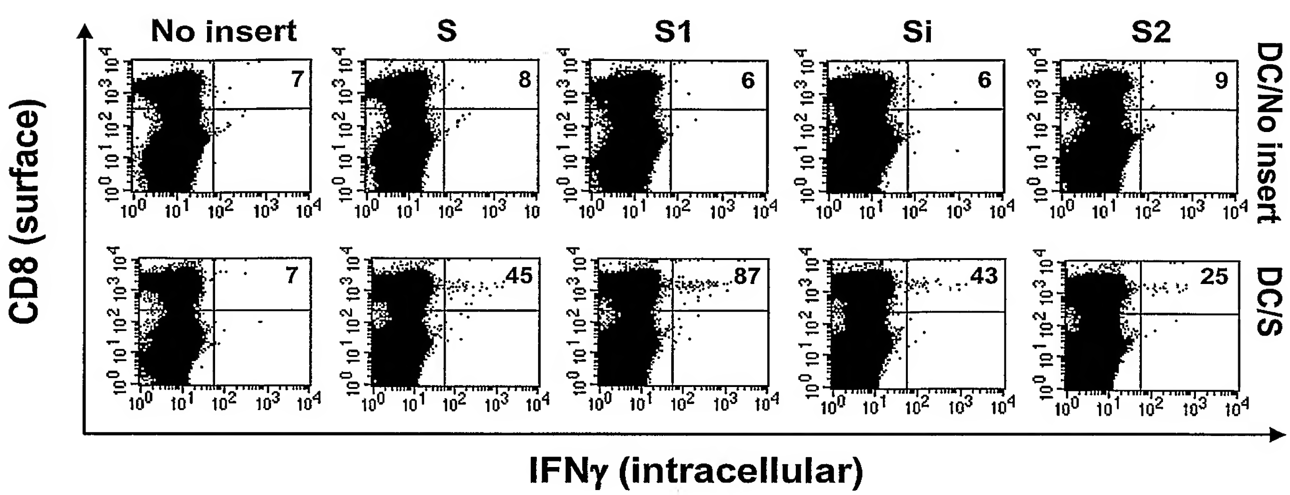

10^5
10^4
10^3
10^2
10^1
10^0
End-point dilution titre
P < 0.003
TC-1/S
TC-1/No insert
No insert
S
S1
Si
S2

Fig. 8B

Fig. 9A


No insert
S
S1
Si
S2
DC/No insert
DC/S
7
8
6
6
9
7
45
87
43
25
CD8 (surface)
IFNγ (intracellular)

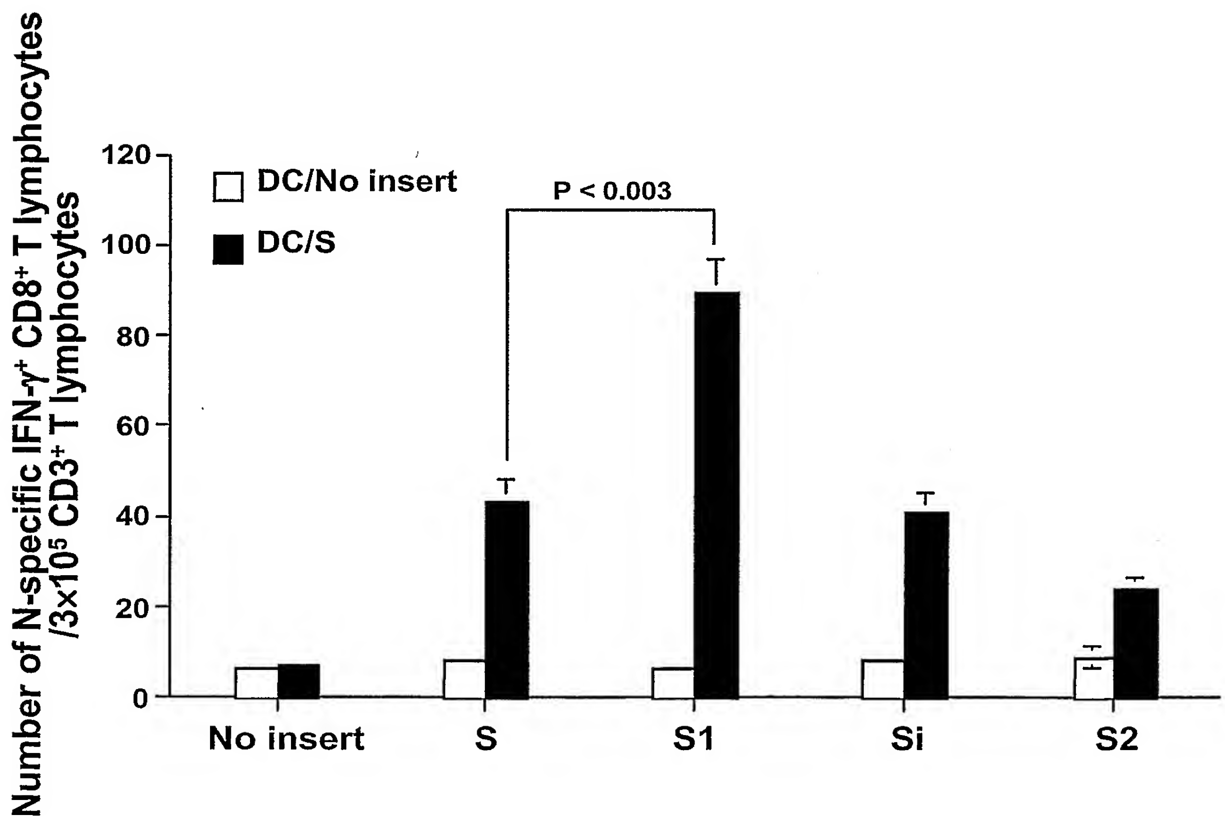

Number of N-specific IFN-γ+ CD8+ T lymphocytes /3x10^5 CD3+ T lymphocytes
DC/No insert
DC/S
P < 0.003
No insert
S
S1
Si
S2

Fig. 9B

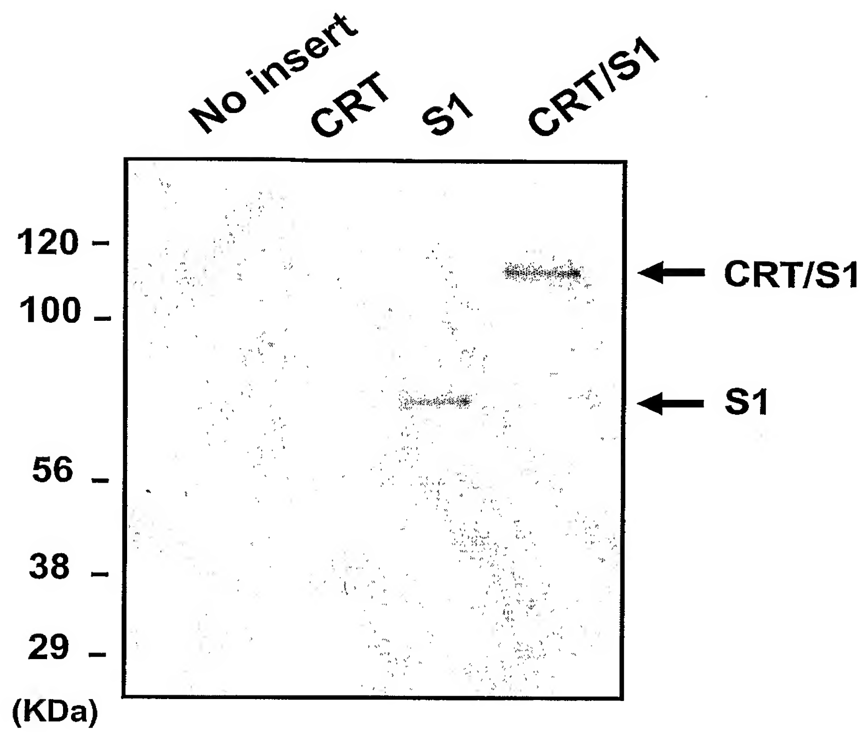


Fig. 10A

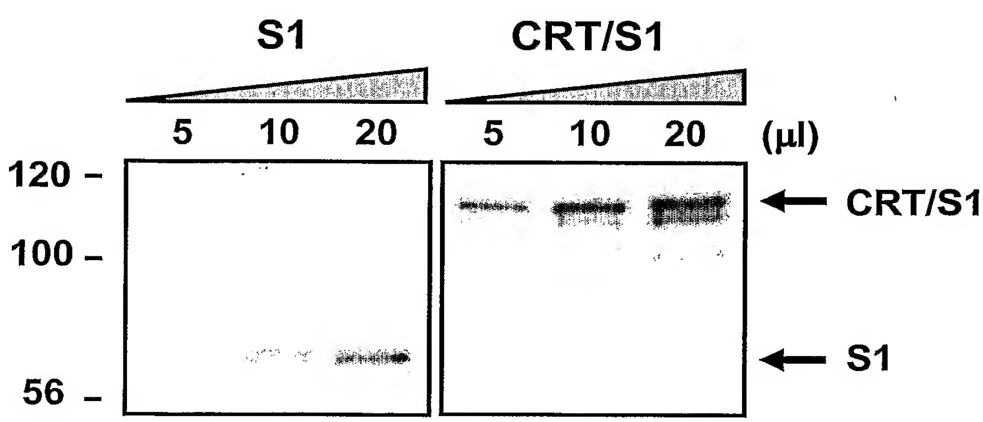


Fig. 10B

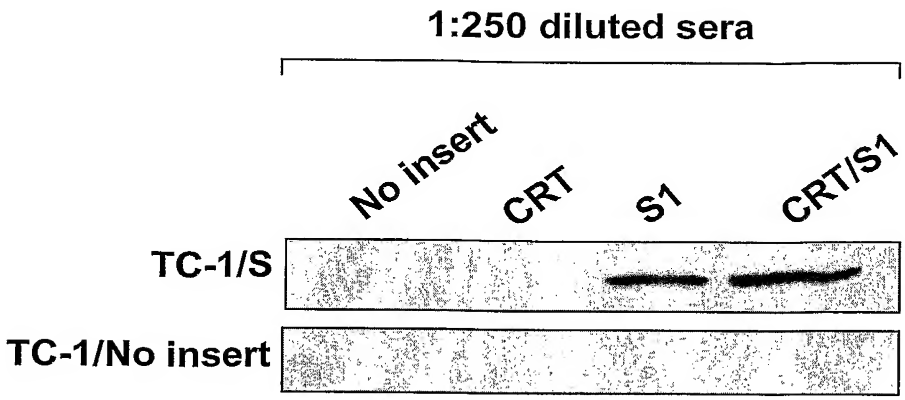


Fig. 11A

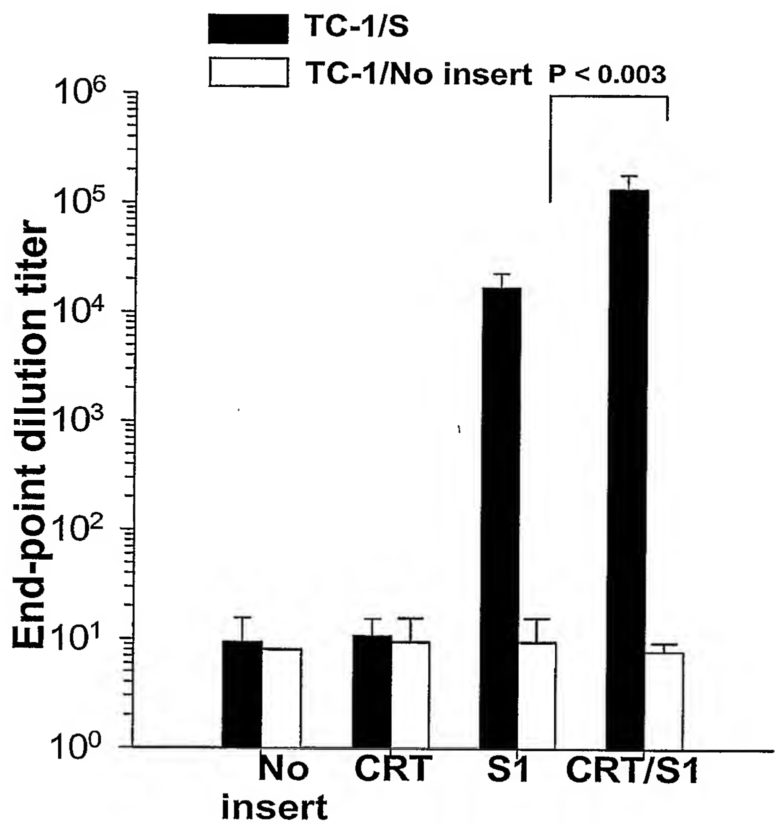
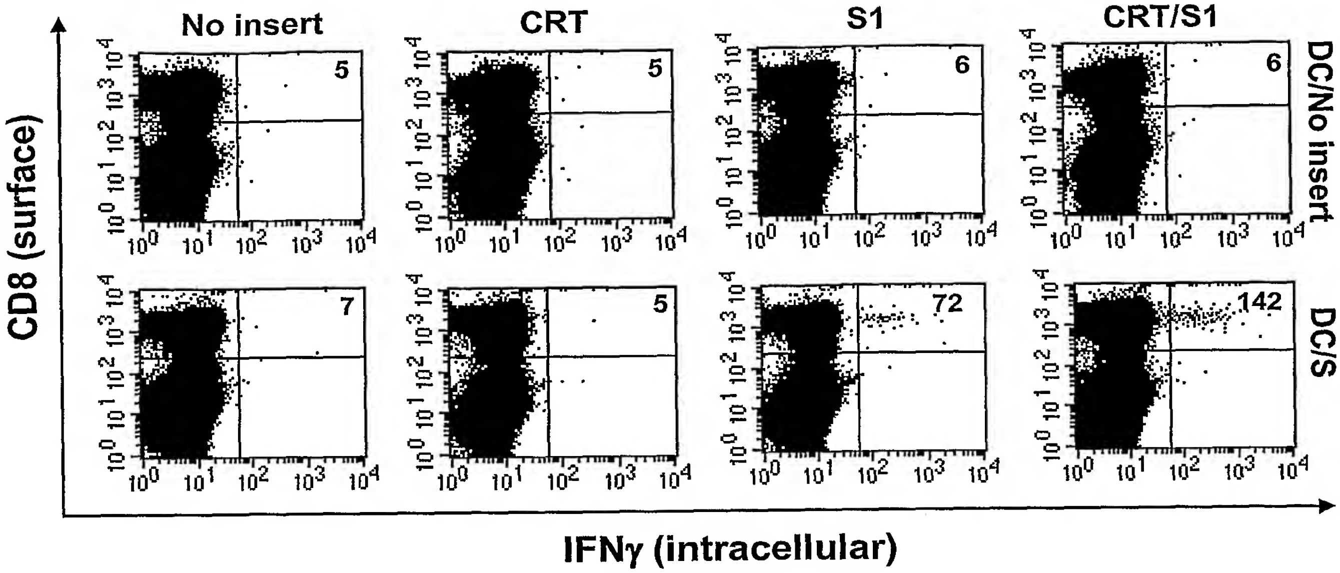


Fig. 11B

Fig. 12A



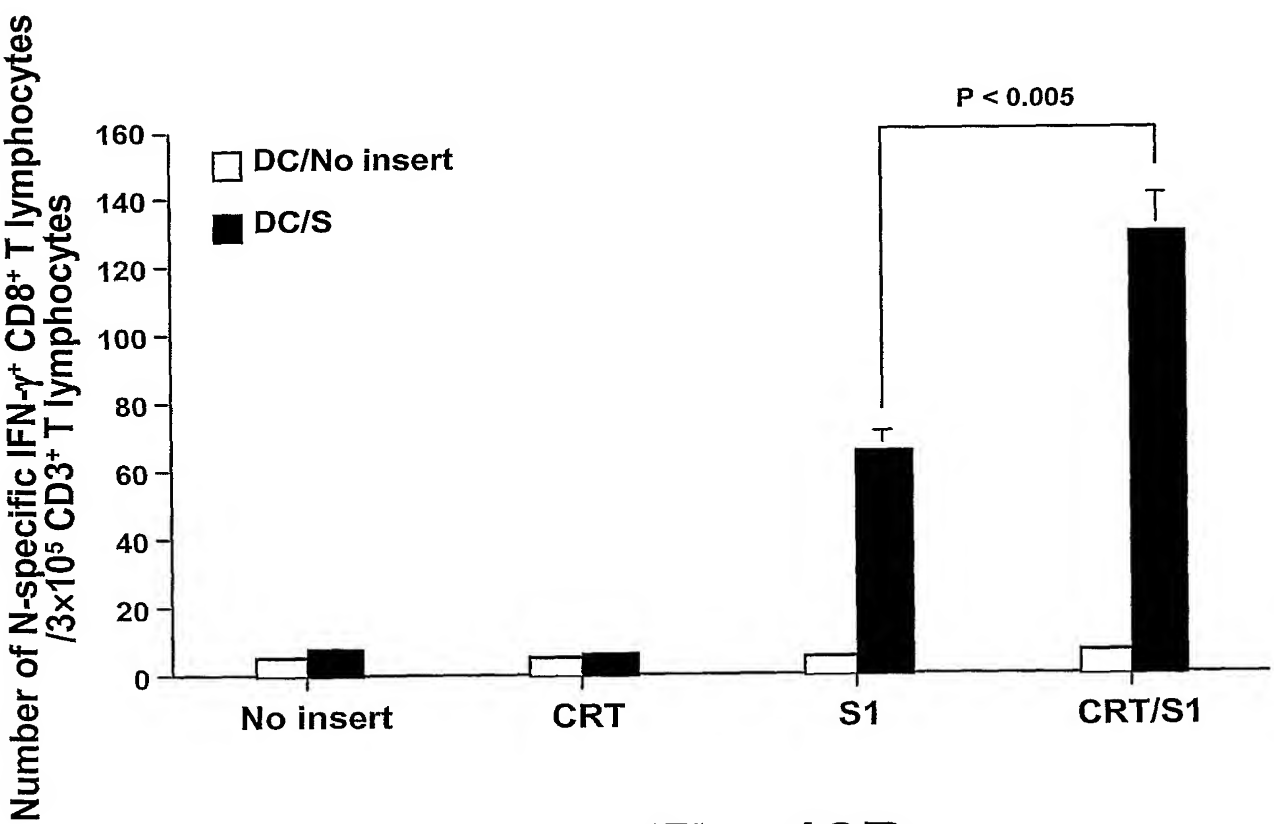


Fig. 12B

Fig. 13A

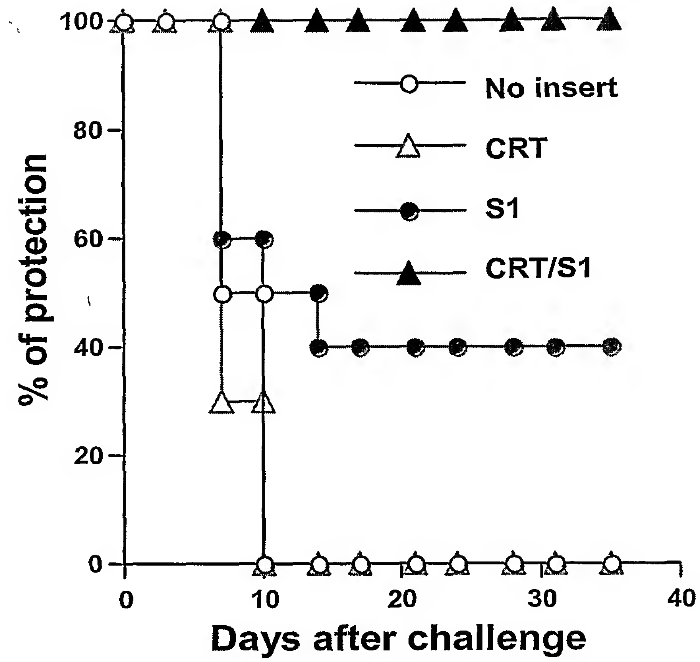
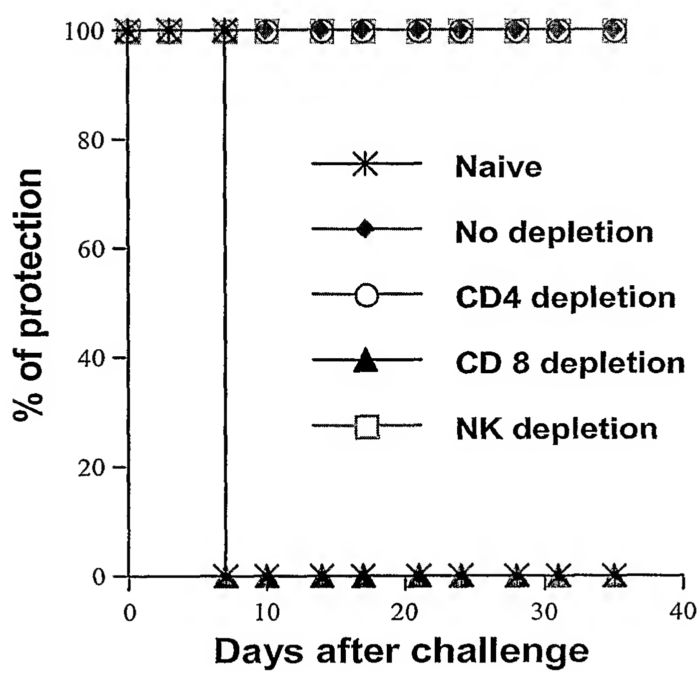


Fig. 13B



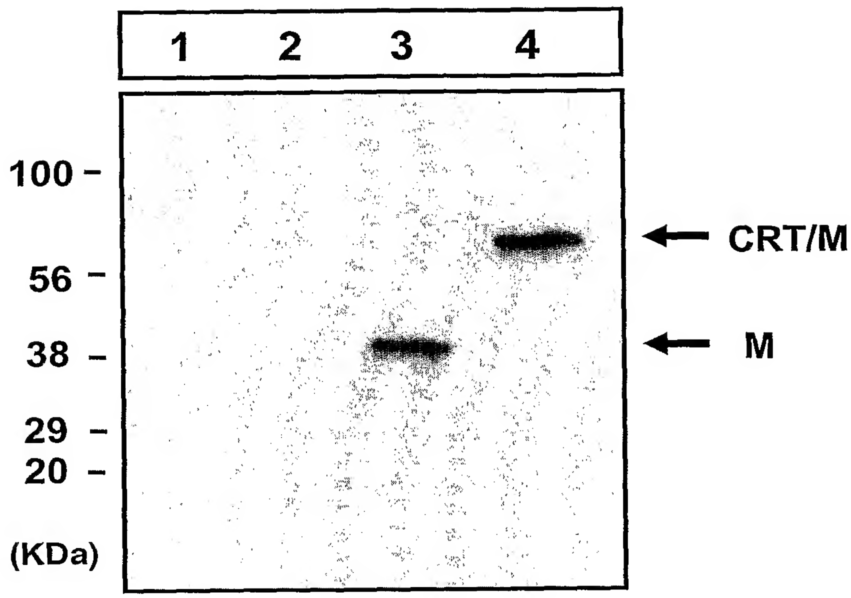
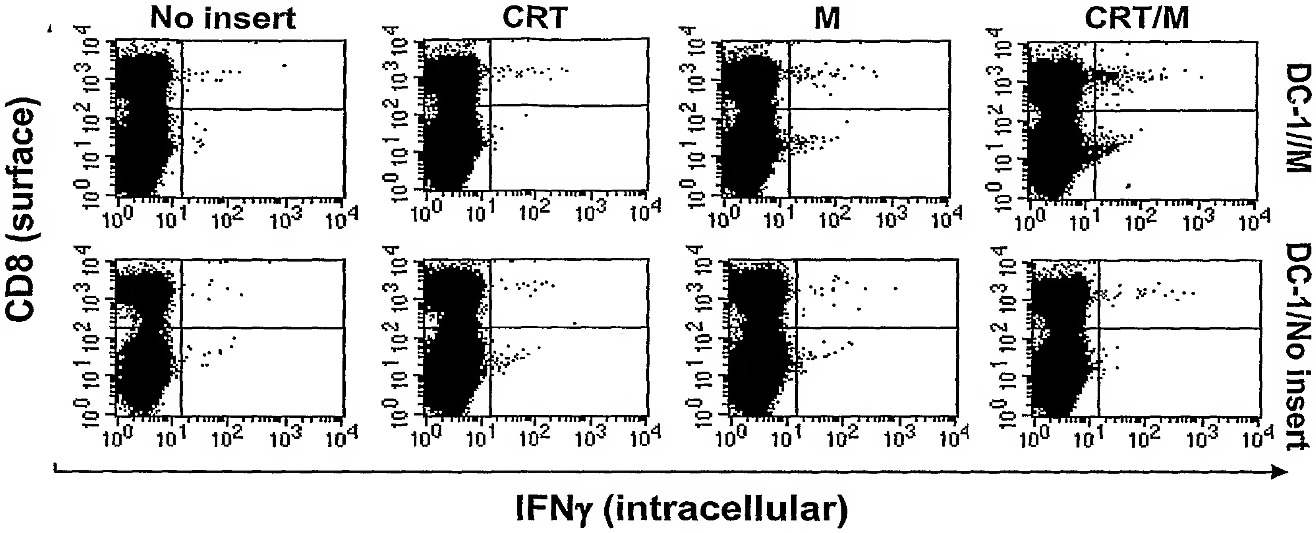


Fig. 14

Fig. 15A



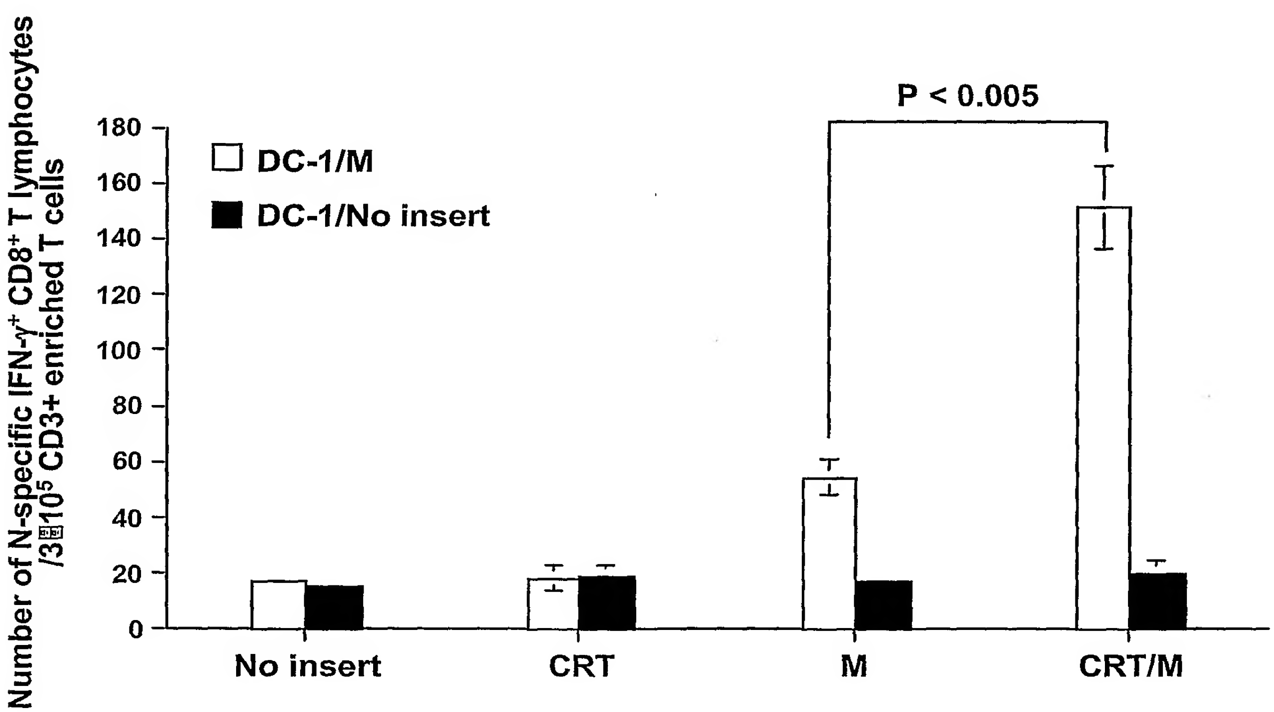


Fig. 15B

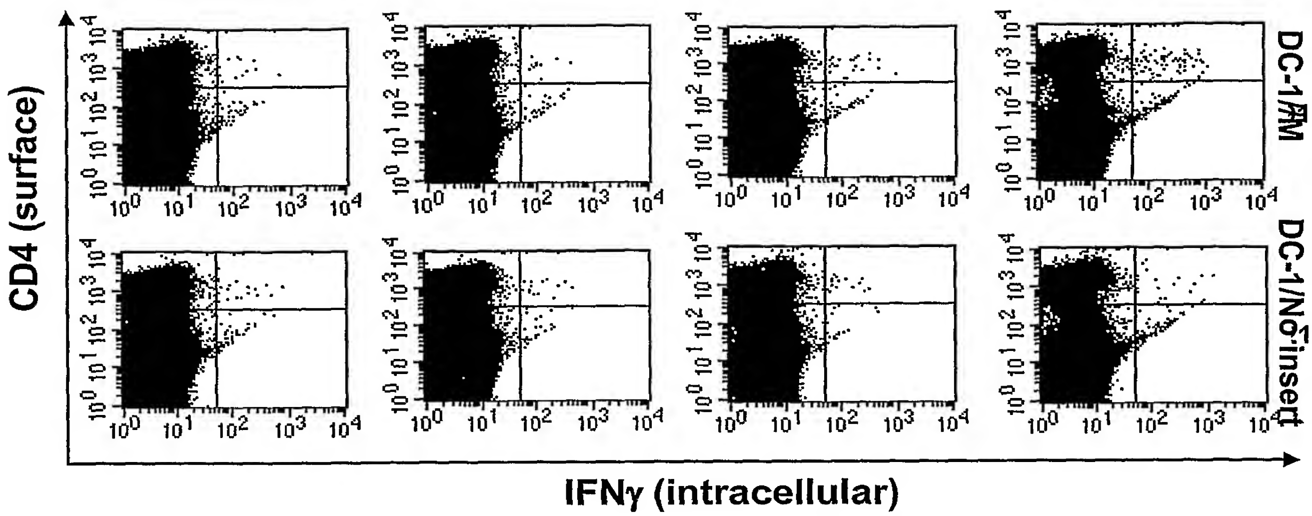


Fig. 16A

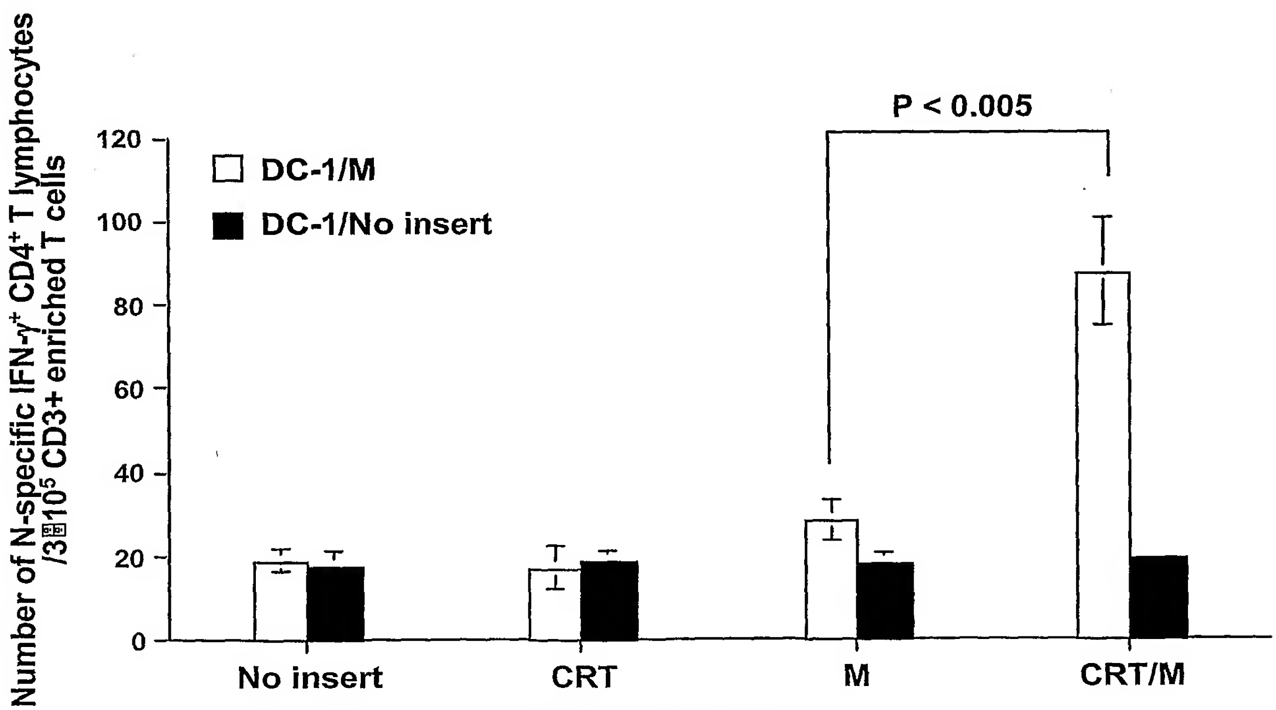


Fig. 16B

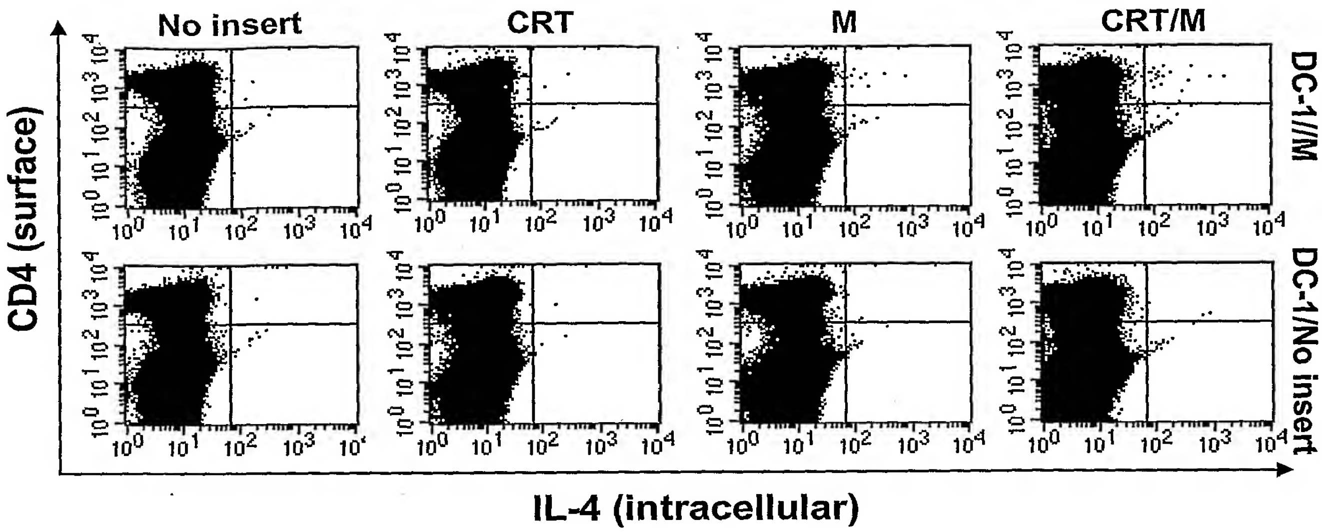


Fig. 17A

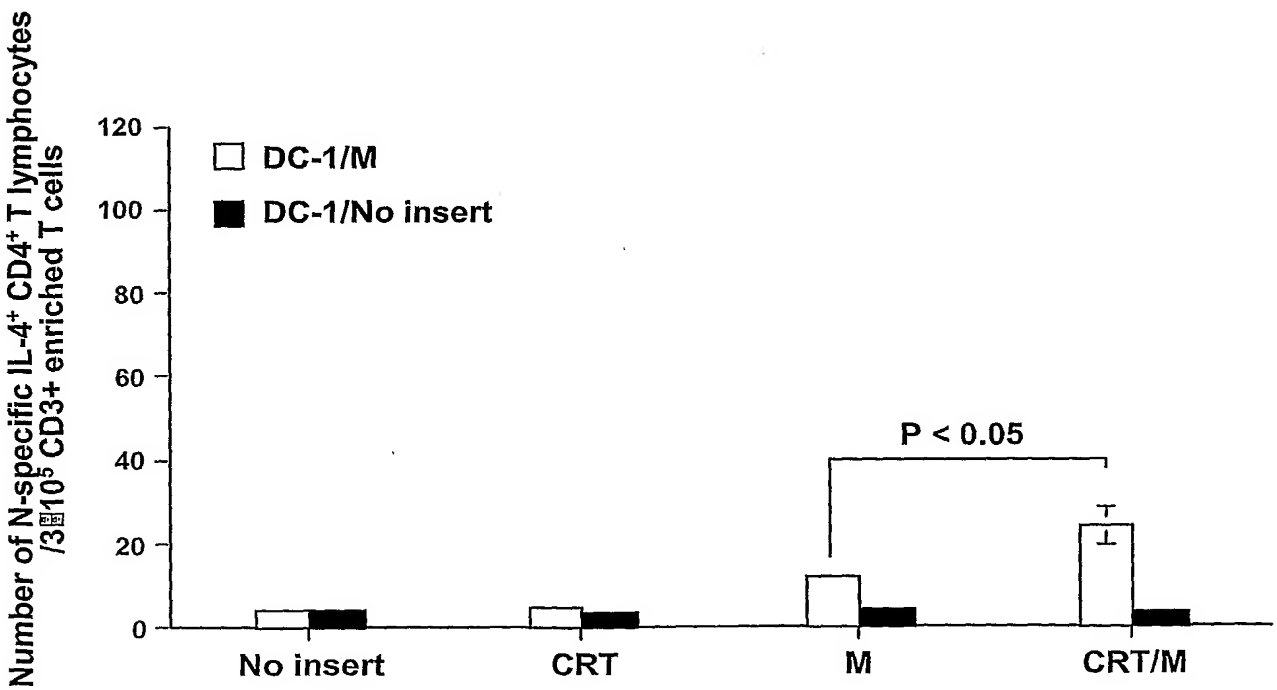


Fig. 17B

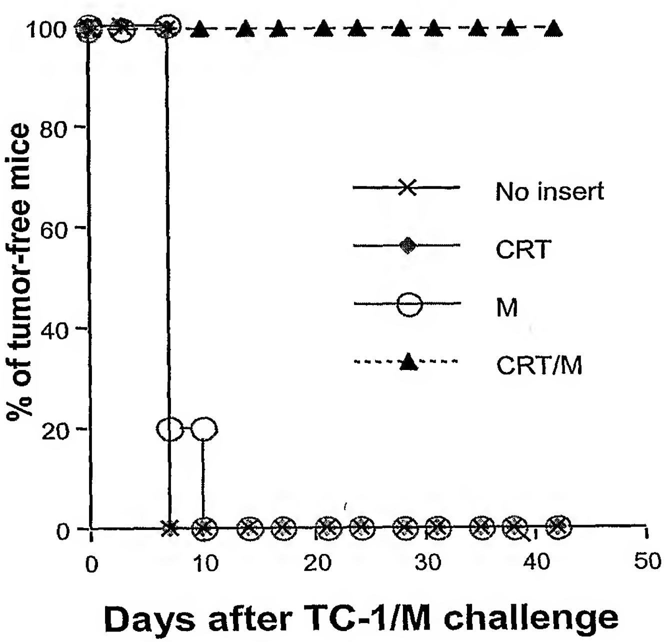


Fig. 18A

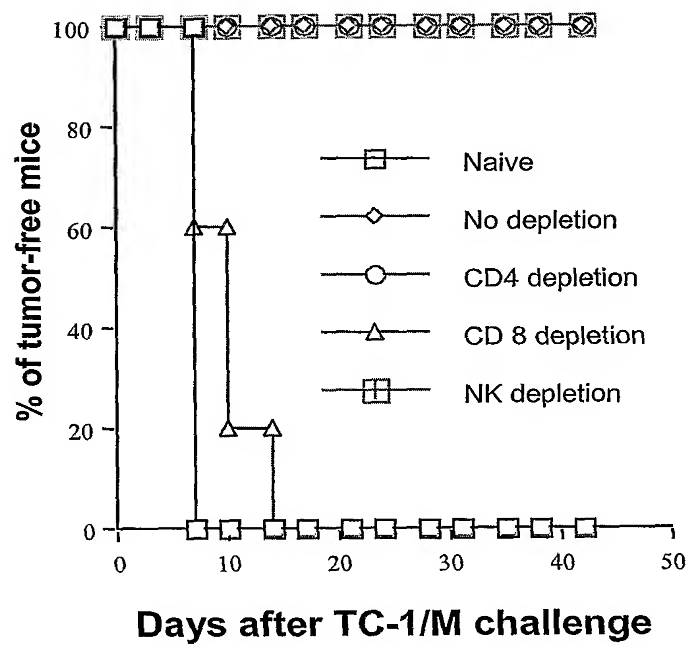
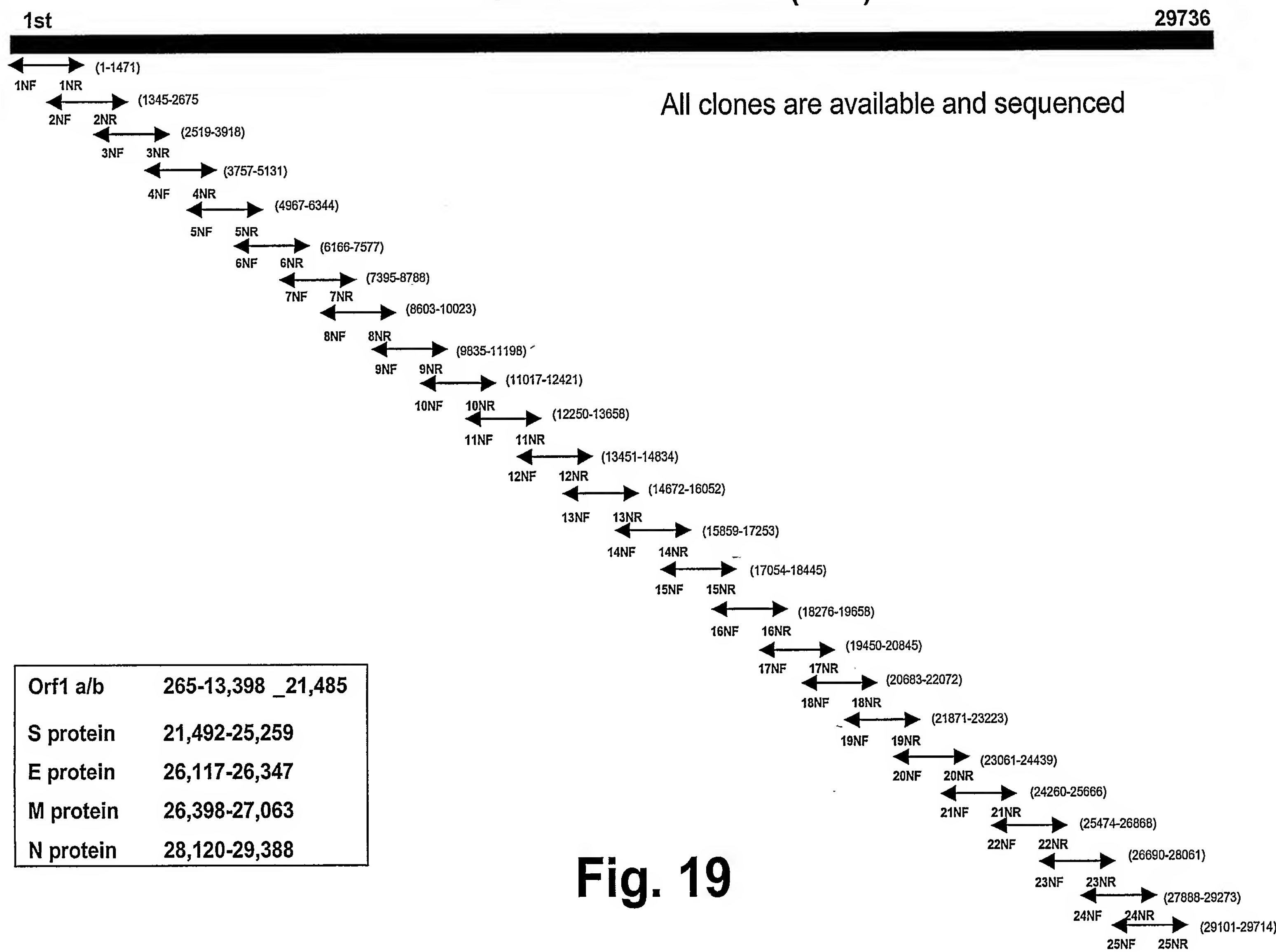


Fig. 18B

SARS Coronavirus (TW1)



Orf1 a/b	265-13,398 _21,485
S protein	21,492-25,259
E protein	26,117-26,347
M protein	26,398-27,063
N protein	28,120-29,388

Fig. 19